STAT 902 - Theory of Probability 2

Typed up by Tuan Hiep Do*

Winter 2024

Taught by Dr. Aukosh Jagannath

Department of Statistics and Actuarial Science,
University of Waterloo, Waterloo, Canada

April 17, 2025

*Electronic address: th2do@uwaterloo.ca

1 Preface

STAT 902 is a graduate level course taught at the University of Waterloo. It is a comprehensive and rigorous introduction to the theory of Brownian motion and stochastic calculus. It is highly recommended that readers should be comfortable with measure theory. A first course on measure theoretic probability will be assumed and some exposure to functional analysis will be helpful but not mandatory.

To get the best out of this course, one should have taken the equivalences of the following courses at the University of Waterloo: PMATH 451 (Measure and Integration), PMATH 453 (Functional Analysis), and STAT 901 (Theory of Probability 1). The author of this course note had only taken measure theory and functional analysis before taking STAT 902. Therefore, one should not be alarmed if they don't have all the desired academic background for this course.

2 Introduction

2.1 Simple Symmetric Random Walk on $\mathbb{Z}$

Let $\{\varepsilon_k\}_{k\in\mathbb{N}}$ be iid random variables such that $\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = -1) = \frac{1}{2}$. Set $S_n = \sum_{i=1}^n \varepsilon_i$. This process fits into two categories for stochastic processes. That is, S_n is a martingale and s_n is a Markov chain with state space $\mathbb{Z}$.

Definition 1. *A Markov chain S_n is recurrent for a state x if $\mathbb{P}(S_n = x) = 1$. The chain is recurrent if it is recurrent for all states.*

Theorem 2. *S_n is recurrent.*

Proof. Note that for all $c \in \mathbb{R}$,

$$\left\{\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} > c\right\}$$

is a tail event with respect to the random variables $\{\varepsilon_i\}_{i\in\mathbb{N}}$. By the Kolmogorov 0-1 law,

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} > c\right) \in \{0,1\}.$$

By the central limit theorem, there exists $\delta_c > 0$ so that for all n sufficiently large, if $\mathbb{P}\left(\frac{S_n}{\sqrt{n}} > c\right) \geq \delta_c$ then

$$\begin{aligned}
\mathbb{P}\left(\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} > c\right) &= \mathbb{P}\left(\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} > c \text{ infinitely often}\right) \\
&= \mathbb{P}\left(\bigcap_{j\geq 1}\bigcup_{n\geq j}\left\{\frac{S_n}{\sqrt{n}} > c\right\}\right) \\
&= \lim_{j\to\infty} \mathbb{P}\left(\bigcup_{n\geq j}\left\{\frac{S_n}{\sqrt{n}} > c\right\}\right) \geq \delta_c > 0.
\end{aligned}$$

And so,

$$\mathbb{P}\left(\limsup_{n\to\infty}\frac{S_n}{\sqrt{n}}=\infty\right)=1$$

whence $\mathbb{P}\left(\limsup_{n\to\infty} S_n=\infty\right)=1$. Similarly, $\mathbb{P}\left(\liminf_{n\to\infty} S_n=-\infty\right)=1$. And so, S_n visits all states with probability 1. $\square$

Other facts that govern S_n are the central limit theorem $\frac{S_n}{\sqrt{n}}\xrightarrow{D}N(0,1)$ and strong law of large number $\frac{S_n}{n}\xrightarrow{a.s.}0$. There is also the law of iterated logarithm which states that $\limsup_{n\to\infty}\frac{|S_n|}{\sqrt{2n\log\log n}}=1$ a.s.

An interesting question one could ask is how long does S_n spend above 0? To answer this, let

$$T_n=|\{k\in\{1,\cdots,n\}:S_k>0|$$

which represents the number of times the walk is above 0 up to time n. In particular, $\frac{T_n}{n}\in[0,1]$.

Lemma 3. $\mathbb{P}(S_n=x)=|$*paths connecting (0, 0) to (n, x)*$|\cdot 2^{-n}$.

Proof. For each path connecting $(0,0)$ to (n,x), the probability of the path occurring is 2^{-n} since

$$\mathbb{P}(\varepsilon_i=1)=\mathbb{P}(\varepsilon_i=-1)=\frac{1}{2}.$$

The lemma then follows. $\square$

Let $u_{2n}=\binom{2n}{n}\cdot 2^{-2n}$ and $f_{2n}=u_{2n-2}-u_{2n}=\frac{1}{2n}u_{2n-2}$.

Lemma 4.

$$\begin{aligned}
u_{2n}&\overset{(1)}{=}\mathbb{P}(S_{2n}=0)\overset{(2)}{=}\mathbb{P}(S_1\neq 0,\cdots,S_{2n}\neq 0)\overset{(3)}{=}\mathbb{P}(S_1\geq 0,\cdots,S_{2n}\geq 0)\\
f_{2n}&\overset{(4)}{=}\mathbb{P}(S_1\neq 0,\cdots,S_{2n-1}\neq 0,S_{2n}=0)\overset{(5)}{=}\mathbb{P}(S_1\geq 0,\cdots,S_{2n-2}\geq 0,S_{2n-1}<0).
\end{aligned}$$

Proof. (1) is clear (equal ups and downs). (4) represents the number of paths from $(0,0)$ to $(2n,0)$ such that $S_1>0,\cdots,S_{2n-1}>0$. Note that

$$\binom{2n-3}{n-1}-\binom{2n-3}{n}=\frac{1}{n}\binom{2n-2}{n-1}.$$

$\binom{2n-3}{n-1}$ is the number of paths from $n=2$ to $2n-1$ and $\binom{2n-3}{n}$ is the number of paths that touch or cross the x-axis. As such, $\frac{1}{n}\binom{2n-2}{n-1}$ is the number of paths from $n=2$ to $2n-1$ that are above the x-axis. However, the reflection principle states that the number of paths touching or crossing the x-axis that go from (x_1,y_1) to (x_2,y_2) is equal to the number of paths that go from $(x_1,-y_1)$ to (x_2,y_2). Thus, the number of paths such that $S_1\neq 0,\cdots,S_{2n-1}\neq 0,S_{2n}=0$ is $\frac{2}{n}\binom{2n-2}{n-1}$. And so,

$$\mathbb{P}(S_1\neq 0,\cdots,S_{2n-1}\neq 0,S_{2n}=0)=\frac{2}{n}\binom{2n-2}{n-1}\cdot 2^{-2n}=f_{2n}.$$

This proves (4). To prove (2), notice that

$$\begin{aligned}\mathbb{P}(S_1 \neq 0, \cdots, S_{2n} \neq 0) &= 1 - f_2 - f_4 - \cdots - f_{2n} \\ &= 1 - (1-u_2) - (u_2 - u_4) - \cdots - (u_{2n-2} - u_{2n}) \\ &= u_{2n}.\end{aligned}$$

The rest is similar. $\square$

Theorem 5. *Let $P_{2k,2n}$ be the probability that the polygonal path connecting $(0,0)$, $(1, S_1), \cdots,$ $(2n, S_{2n})$ is above 0 for $2k$ units and below 0 for $2n-2k$ units. Then $P_{2k,2n} = u_{2k} \cdot u_{2n-2k}$.*

Proof. By symmetry, $P_{2n,2n} = u_{2n} = P_{0,2n}$. Let $1 \leq k \leq n-1$ then observe that for such a k, the path must cross the x-axis at some point. Let $2r$ denote such crossing. Then either

1. The path is initially positive and then on the interval $(2r, 2n)$, it spends $2k-2r$ units positive, $2n-2k$ units negative. Therefore,

$$\text{number of paths} = 2^{2r} \cdot \frac{f_{2r}}{2} \cdot 2^{2n-2r} \cdot P_{2k-2r,2n-2r}.$$

2. The path is initially negative and then spends $2k$ units positive, $2n-2k-2r$ units negative. Thus,

$$\text{number of paths} = 2^{2r} \cdot \frac{f_{2r}}{2} \cdot 2^{2n-2r} \cdot P_{2k,2n-2r}.$$

Adding over r and multiply by 2^{-2n} to get

$$P_{2k,2r} = \frac{1}{2}\sum_{r=1}^{k} f_{2r} \cdot P_{2k-2r,2n-2r} + \frac{1}{2}\sum_{r=1}^{n-k} f_{2r} \cdot P_{2k,2n-2r}.$$

By induction, $P_{2k,2r} = u_{2k} \cdot u_{2r-2k}$ for $r \leq n-1$. And so,

$$P_{2k,2n} = \frac{1}{2} u_{2n-2k} \underbrace{\sum_{r=1}^{k} f_{2r} \cdot u_{2k-2r}}_{u_{2k}} + \frac{1}{2} u_{2k} \underbrace{\sum_{r=1}^{n-k} f_{2r} u_{2n-2r-2k}}_{u_{2n-2k}}.$$

$\square$

Theorem 6. *(Arcsine law for simple random walk)*

$$\mathbb{P}\left(\frac{T_n}{n} \leq \alpha\right) \overset{n\to\infty}{\to} \frac{2}{\pi}\arcsin(\alpha^{1/2}).$$

Proof. Recall Stirling's formula which states that

$$n! \sim (2\pi)^{1/2} n^{n+1/2} e^{-n}.$$

And so, for sufficiently large k and $n-k$,

$$P_{2k,2n} \sim \frac{1}{\pi k^{1/2}(n-k)^{1/2}}.$$

Hence if $\alpha \in (0,1)$,

$$\mathbb{P}\left(\frac{1}{2} \le \frac{T_n}{n} \le \alpha\right) = \sum_{\frac{n}{2} \le k \le \alpha n} P_{2k,2n} \sim \frac{1}{\pi n} \sum_{\frac{n}{2} \le k \le \alpha n} \left[\frac{k}{n}\left(1 - \frac{k}{n}\right)\right]^{-1/2}$$
$$\sim \frac{1}{\pi} \int_{\frac{1}{2}}^{\alpha} \frac{1}{[x(1-x)]^{1/2}}\, dx$$
$$= \frac{2}{\pi} \arcsin(\alpha^{1/2}) - \frac{1}{2}.$$

By symmetry, $\mathbb{P}(\frac{T_n}{n} \le \frac{1}{2}) \to \frac{1}{2}$ as $n \to \infty$. Substituting into $\mathbb{P}(\frac{T_n}{n} \le \alpha) = \mathbb{P}\left(\frac{1}{2} \le \frac{T_n}{n} \le \alpha\right) - \mathbb{P}(\frac{T_n}{n} \le \frac{1}{2})$, the result follows. $\square$

2.2 Ito's Formula for Simple Random Walk

Let $S_n = \sum_{j=1}^{n} \varepsilon_j$ where $\{\varepsilon_j\}_{j \in \mathbb{N}}$ are iid random variables with $\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = -1) = \frac{1}{2}$. We have seen that S_n is a recurrent Markov chain on $\mathbb{Z}$. The amount of time T_n that S_n spends above 0 follows arcsine law

$$\mathbb{P}\left(\frac{T_n}{n} \le \alpha\right) \overset{n\to\infty}{\to} \frac{2}{\pi} \arcsin(\alpha^{1/2}).$$

Note that S_n is a martingale with respect to $\mathcal{F}_n = \sigma(\varepsilon_i : 1 \le i \le n)$ in a probability space triplet $(\Omega, \mathcal{F}, \mathbb{P})$. That is, $\mathbb{E}(S_n|\mathcal{F}_{n-1}) = S_{n-1}$. Also, $\mathcal{F}_n$ only has finitely many sets and all sets are generated by

$$A_n^i = \{\varepsilon_1 = a_i^{(1)}, \cdots, \varepsilon_n = a_i^{(n)}\}$$

for $1 \le i \le 2^n$ and $a_i^{(j)} \in \{-1, 1\}$. If X is $\mathcal{F}_n$-measurable random variable then $X = \sum_{i=1}^{2^n} X_i \mathbb{1}_{A_n^i}$.

Theorem 7. *Suppose that m_n is a martingale adapted to $\mathcal{F}_n$. Then m_n can be uniquely represented as a martingale transform using S_n as*

$$m_n = m_0 + \sum_{k=1}^{n-1} (S_{k+1} - S_k) \cdot \xi_k$$

where ξ_k is $\mathcal{F}_k$-measurable.

Proof. By telescoping sum,

$$m_n = m_0 + \sum_{k=1}^{n-1} \underbrace{\frac{m_{k+1} - m_k}{S_{k+1} - S_k}}_{\xi_k} (S_{k+1} - S_k).$$

We now show that ξ_k is $\mathcal{F}_k$-measurable. Notice that each set generating $\mathcal{F}_{k+1}$ can be written as

$$A_k^{i,1} = A_k^i \cap \{\varepsilon_{k+1} = 1\} \text{ or } A_k^{i,-1} = A_k^i \cap \{\varepsilon_{k+1} = -1\}.$$

Then

$$m_k = \sum_{j=1}^{2^k} m_k^j \mathbb{1}_{A_k^i},$$

$$m_{K+1} = \sum_{j=1}^{2^k} \left(m_{k+1}^{j,1} \mathbb{1}_{A_k^{j,1}} + m_{k+1}^{j,-1} \mathbb{1}_{A_k^{j,-1}} \right).$$

Note that $\mathbb{1}_{A_k^{j,1}} = \mathbb{1}_{A_k^j} \mathbb{1}_{\{\varepsilon_{k+1}=1\}}$ and $\mathbb{1}_{A_k^{j,-1}} = \mathbb{1}_{A_k^j} \mathbb{1}_{\{\varepsilon_{k+1}=-1\}}$. This in conjunction with $\mathbb{E}(m_{k+1}|\mathcal{F}_k) = m_k$ implies that

$$m_k^j = m_{k+1}^{j,1} \mathbb{E}[\mathbb{1}_{\{\varepsilon_k=1\}}|\mathcal{F}_k] + m_{k+1}^{j,-1} \mathbb{E}[\mathbb{1}_{\{\varepsilon_k=-1\}}|\mathcal{F}_k] = \frac{m_{k+1}^{j,1} + m_{k+1}^{j,-1}}{2}.$$

On the set $A_k^{j,1}$, $\varepsilon_k = m_{k+1}^{j,1} - m_k^j = \frac{m_{k+1}^{j,1} - m_{k+1}^{j,-1}}{2}$ and on the set $A_k^{j,-1}$,

$$\xi_k = \frac{m_{k+1}^{j,-1} - m_k^j}{-1} = m_k^j - m_{k+1}^{j,-1} = \frac{m_{k+1}^{j,1} - m_{k+1}^{j,-1}}{2}.$$

In both cases, ξ_k is a function of $\varepsilon_1, \cdots, \varepsilon_k$ whence it is $\mathcal{F}_k$-measurable. Uniqueness is clear. $\square$

Consider any $f : \mathbb{Z} \to \mathbb{R}$. The process $f(S_n)$ satisfies the Doob Decomposition, namely $f(S_n) = m_n + A_n$ where m_n is a martingale, A_n is $\mathcal{F}_{n-1}$-measurable, $A_0 = 0$, and

$$A_{k+1} - A_k = \mathbb{E}[f(S_{k+1}) - f(S_k)|\mathcal{F}_k],$$
$$m_0 = f(S_0),$$
$$m_{k+1} - m_k = f(S_{k+1}) - \mathbb{E}[f(S_{k+1})|\mathcal{F}_k].$$

Define $f'_+(a) = f(a+1) - f(a)$ and $f'_-(a) = f(a) - f(a-1)$. Let

$$f'(a) = \frac{f'_+(a) + f'_-(a)}{2}$$
$$f''(a) = f'_+(a) - f'_-(a) = f(a+1) + f(a-1) - 2f(a).$$

These are the discrete analogs of the derivatives of f. Then,

$$A_{k+1} - A_k = \mathbb{E}[f(S_{k+1}) - f(S_k)|\mathcal{F}_k] = \frac{f(S_k+1)}{2} + \frac{f(S_k-1)}{2} - f(S_k) = \frac{1}{2} f''(S_k).$$

Consider the martingale transform of m_n using S_n

$$m_n = m_0 + \sum_{k=1}^{n-1} \xi_k \cdot (S_{k+1} - S_k)$$

where

$$\xi_k = \frac{m_{k+1} - m_k}{S_{k+1} - S_k} = \frac{f(S_{k+1}) - \mathbb{E}[f(S_{k+1})|\mathcal{F}_k]}{S_{k+1} - S_k} = \frac{f(S_{k+1}) - \frac{f(S_k+1)}{2} - \frac{f(S_k-1)}{2}}{\varepsilon_k}.$$

When $\varepsilon_k = 1$ then $\xi_k = f(S_{k+1}) - \frac{f(S_k+1)}{2} - \frac{f(S_k-1)}{2} = f'(S_k)$. This also holds when $\varepsilon_k = -1$. Using the fact that $f(S_n) = m_n + A_n$, the following results follows readily

Theorem 8. *(Ito's formula for simple random walk)*
For any $f : \mathbb{Z} \to \mathbb{R}$,

$$f(S_n) = f(S_0) + \sum_{k=1}^{n-1} f'(S_k)(S_{k+1} - S_k) + \frac{1}{2}\sum_{k=1}^{n-1} f''(S_k)$$

where S_n denotes a simple random walk.

A more general result can be restated as follows. Denote the local time of a random walk S_n to be a function $L : \mathbb{N} \times \mathbb{Z} \to \mathbb{N}$

$$L(n, a) = \sum_{k=0}^{n-1} \mathbb{1}_{\{S_k = a\}}.$$

This represents how much time the walk spends at a. The alternative Ito's formula for simple random walk is

$$f(S_n) = f(S_0) + \sum_{k=1}^{n-1} f'(S_k)(S_{k+1} - S_k) + \frac{1}{2}\sum_{z \in \mathbb{Z}} f''(z) \cdot L(n, z).$$

3 Brownian Motion

3.1 Motivation

As readers have seen in previous sections, any function $f : \mathbb{Z} \to \mathbb{R}$ can be decomposed into the following form using the simple random walk S_n

$$f(S_n) = f(S_0) + \sum_{k=1}^{n-1} f'(S_k)(S_{k+1} - S_k) + \text{martingale}.$$

The question is can it be generalized to any sufficiently "nice" functions on $\mathbb{R}$? What if instead of $\varepsilon_i \sim \text{Bernoulli}(\frac{1}{2})$, one only has that ε_i are iid with $\mathbb{E}(\varepsilon_i) = 0$ and $\mathbb{E}(\varepsilon_i^2) < \infty$? How about answering questions regarding limits such as

$$S_n^* = \max_{k \le n} S_k \to ?$$
$$\frac{T_n}{n} = \frac{|\{i > 0 : S_i > 0\}}{n} \to ?$$

That is, we want to build a universal scaling limit W_t of $S_{n(t)}$ so that

$$W_t = \delta_t S_{n(t)} \sim N(0, \delta_t^2 \cdot n(t)) \sim N(0, t).$$

Observe that $S_n - S_m \perp\!\!\!\perp S_m - S_k$ for $k < m < n$ ($\perp\!\!\!\perp$ denotes independence) so we would also want that $W_{t_2} - W_{t_1} \perp\!\!\!\perp W_{t_1} - W_{t_0}$ for $t_0 < t_1 < t_2$ (independent increments). This leads to the following definition

Definition 9. *A real-valued process $(B_t)_{t\geq 0}$ indexed by $\mathbb{R}_+ = [0,\infty)$ is a Brownian motion with initial data $x \in \mathbb{R}$ if*

1. *$B_0 = x$,*

2. *(B_t) has independent increments,*

3. *$B_{t+h} - B_t \stackrel{D}{=} N(0,h)$,*

4. *The map $t \mapsto B_t$ is continuous a.s.*

With this abstract definition, the questions now are

1. Does Brownian motion exist?

2. In what sense is it a scaling limit?

We explore each of these questions in the following sections.

3.2 Construction

In this section, three separate constructions of Brownian motions are presented using pure measure theory, Wiener construction, and Levy construction respectively.

3.2.1 Attempt 1: Pure Measure Theory

Consider the space of $\mathbb{R}^{[0,\infty)} = \{f : \mathbb{R}_+ \to \mathbb{R}\}$ equipped with the product σ-algebra.

Definition 10. *Let $T \subseteq \mathbb{R}$ be an interval. For all $t_1, \cdots, t_k \in T$, we say that $\nu_{t_1,\cdots,t_k} \in M_1(\mathbb{R}^k)$—the space of probability measures on $\mathbb{R}^k$—are consistent if*

1. *For any permutation σ of $\{1,\cdots,k\}$ and measurable $F_1,\cdots,F_k$,*

$$\nu_{t_1,\cdots,t_k}(F_1 \times \cdots \times F_k) = \nu_{t_{\sigma(1)},\cdots,t_{\sigma(k)}}(F_{\sigma(1)} \times \cdots \times F_{\sigma(k)}).$$

2. *For all measurable $F_1,\cdots,F_k$ and $m \geq 1$,*

$$\nu_{t_1,\cdots,t_k}(F_1 \times \cdots \times F_k) = \nu_{t_1,\cdots,t_k,t_{k+1},\cdots,t_{k+m}}(F_1 \times \cdots \times F_k \times \mathbb{R}^m).$$

Theorem 11. *(Kolmogorov Extension Theorem)*
Given an interval $T \subseteq \mathbb{R}$ and a collection of consistent measures $(\nu_{t_1,\cdots,t_k})$, there exists a measure $\nu \in M_1((\mathbb{R}^T, \mathscr{B}_{prod}))$ whose marginals are $(\nu_{t_1,\cdots,t_k})$.

With the measure ν as above, one can define a stochastic process $(B_t)_{t\geq 0}$ with the desired properties 1, 2, and 3 in the definition 9 of Brownian motion. Such a process is called a *pre-Brownian motion*. However, the event $\{f \text{ is continuous}\}$ is not $\mathscr{B}_{\text{prod}}$-measurable so property 4 is not satisfied. To remedy the situation, a modification of such a process is constructed instead.

Definition 12. *Given X, $\tilde{X}$ is called a modification of X if $\mathbb{P}(\tilde{X}_t = X_t) = 1$ for all $t \in T$.*

Theorem 13. *(Kolmogorov Continuity Criteria)*
Let $(X_t)_{t\in I}$ (I is a bounded interval) with values in a complete separable metric space $(\mathcal{X}, d)$. Suppose that there exists $\alpha, \beta, k > 0$ with

$$\mathbb{E}(d(X_t, X_s)^\alpha) \leq k|t-s|^{1+\beta}$$

for all $t, s \in I$. Then there exists a modification $\tilde{X}$ of X such that $\tilde{X}$ is γ-Holder continuous where $\gamma \in (0, \frac{\beta}{\alpha})$. That is, $|\tilde{X}_t - \tilde{X}_s| \leq L|t-s|^\gamma$ for some L.

Using the extension theorem and continuity criteria, one can prove the existence of Brownian motion on $\mathbb{R}^d$. This was given as a homework problem.

3.2.2 Attempt 2: Wiener

The idea here is to use the Kolmogorov Extension Theorem to construct $\underline{\mathbf{g}} = (g_1, g_2, \cdots) \in \mathbb{R}^\infty$ which are iid Gaussians. One will view $\underline{\mathbf{g}}$ as a Fourier transform of $f(t) = \sum g_k \cos(2\pi kt)$. It is worth noting that f is not a function but rather a distribution. From here, let $f = \dot{B}_t = \frac{d}{dt}B_t$ then B_t can be found by integrating f.

Recall that

$$L^2([0,1], dx) = \left\{ f : [0,1] \to \mathbb{C} : \int_0^1 |f|^2 \, dx < \infty \right\}$$

is a Hilbert space equipped with the inner product $\langle f, g\rangle = \int_0^1 f\overline{g}\, dx$. Additionally, the space

$$\ell_2 = \left\{ \underline{\mathbf{a}} = (a_1, a_2, \cdots) : \sum |a_i|^2 < \infty \right\}$$

is also a Hilbert space equipped with the inner proudct $\langle \underline{\mathbf{a}}, \underline{\mathbf{b}} \rangle = \sum a_i \overline{b}_i$. In particular, the following result was obtained from previous real analysis course.

Theorem 14. *(Fourier)*
L^2 is isometrically isomorphic to ℓ_2. That is, there exists an invertible linear map $T : L^2 \to \ell_2$ so that $\langle f, g\rangle_{L^2} = \langle Tf, Tg\rangle_{\ell_2}$.

Definition 15. *For $f \in L^2$, the Fourier transform of f is $\hat{f} : \mathbb{Z} \to \mathbb{R}$ such that $\hat{f}(k) = \int_0^1 f(x)e^{2\pi ikx}\, dx$.*

Note that for smooth f, $\hat{f}'(k) = k\hat{f}(k)$. If $g(t) = \int_0^t f\, dx$ then $\hat{g}(k) = \frac{\hat{f}(k)}{k}$. Let $\hat{h}(k) = \frac{g_k}{k}$ then

$$\mathrm{Re}(h(x)) = \sum \frac{g_k}{k} \cos(2\pi kx).$$

Lemma 16. $\|h\|_{L^2}^2 = \left\|\hat{h}\right\|_{\ell_2}^2 < \infty$.

Proof. Observe that since $\mathbf{g}$ are iid Gaussians,

$$\sum \frac{\mathbb{E}(g_k^2)}{k^2} = \mathbb{E}(g_1^2) \sum \frac{1}{k^2} < \infty,$$
$$\sum \frac{\mathbb{E}(g_k^4)}{k^4} = \mathbb{E}(g_1^4) \sum \frac{1}{k^4} < \infty.$$

The result then follows from the Kolmogorov two-series theorem. □

If one denotes

$$B_t = \underbrace{\mathrm{Re}(h(t))}_{\text{Brownian bridge}} + g_0 t = g_0 t + \sum_{k=-\infty}^{\infty} \frac{g(k)}{k} \cos(2\pi k t)$$

then this is a potential candidate. However, the question is whether $B(t)$ is continuous. This can be answered using either the Kolmogorov Continuity Criteria 13 or a lot of grunt work...

Suppose that $\dot{B}_t$ made sense then by Plancherel theorem, $\int f(t)\dot{B}(t)\,dt = \int f(t)\,dB_t$. In particular,

$$\langle f, \dot{B} \rangle_{L^2} = \langle \hat{f}, \hat{\dot{B}} \rangle_{\ell_2} = \sum_{k \in \mathbb{Z}} \hat{f}(k) g_k.$$

Lemma 17. *This series converges a.s.*

Proof. Direct application of Kolmogorov two-series theorem gives the desired conclusion. □

In L^2, let

$$\hat{f}_M = \begin{cases} \hat{f}(k) & |k| \le M, \\ 0 & \text{otherwise.} \end{cases}$$

Then,

$$\mathbb{E}\left[\left(\sum \hat{f}_M(k) g_k - \sum \hat{f}(k) g_k\right)^2\right] = \mathbb{E}\left[\left(\sum_{|k|>M} f(k) g_k\right)^2\right] \le \sum_{|k|>M} |f(k)|^2 \overset{m\to\infty}{\to} 0.$$

We will state an easy result about limiting Gaussians and leave the proof to the readers.

Theorem 18. *Suppose that $Z_n \sim N(m_n, \sigma_n^2)$ are Gaussians. If $Z_n \overset{(D)}{\to} Z$ then $Z \sim N(m, \sigma^2)$ where $\lim m_n = m$ and $\lim \sigma_n^2 = \sigma^2$.*

And so,

$$I(f_M) := \underbrace{\sum_{|k| \le M} \hat{f}(k) g_k}_{N(0, \|f_M\|_{L^2}^2} \overset{M\to\infty}{\to} I(f) := \sum \hat{f}(k) g_k \sim N(0, \|f\|_{L^2}^2).$$

Definition 19. *(Centered Gaussian space)*
Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ admits Gaussian random variables. The set $\mathcal{H} = \{\text{centered Gaussians}\} \subseteq L^2(\Omega, \mathcal{F}, \mathbb{P})$ is called the centered Gaussian space.

Theorem 20. *(Wiener's isometry)*
The mapping $I : L^2([0,1], dx) \to \mathcal{H}$ is a linear isometry.

Proof. Note that

$$\widehat{af + bh}(k) = \int_0^1 (af(x) + bh(x))e^{2\pi i k x}\, dx = a\hat{f}(k) + b\hat{h}(k).$$

And so,

$$\begin{aligned} I(af + bh) = \sum (a\hat{f}(k) + b\hat{h}(k))g_k &= a \sum \hat{f}(k) g_k + b \sum \hat{h}(k) g_k \\ &= aI(f) + bI(g) \end{aligned}$$

implying that I is linear. On the other hand, observe that if $\hat{f}(k), \hat{h}(k) = 0$ for $|k| \geq M$ then

$$\begin{aligned} \langle I(f), I(h)\rangle = \mathbb{E}(I(f)I(h)) = \mathbb{E}\left[\left(\sum \hat{f}(k) g_k\right)\left(\sum \hat{h}(k) g_k\right)\right] &= \sum_{k,\ell} \hat{f}(k)\hat{h}(\ell)\,\underbrace{\mathbb{E}(g_k g_\ell)}_{\delta_{k\ell}} \\ &= \sum \hat{f}(k)\hat{h}(k) \\ &= \langle f, h\rangle_{L^2} \end{aligned}$$

by Plancherel identity. However,

$$\begin{aligned} \langle I(f), I(h)\rangle - \langle I(f_M), I(h_M)\rangle &= \langle I(f) - I(f_M), I(h)\rangle + \langle I(f_M), I(h) - I(h_M)\rangle \\ &\leq \|I(f) - I(f_M)\|_{L^2} \|I(h)\|_{L^2} + \|I(f_M)\|_{L^2} \|I(h) - I(h_M)\|_{L^2}. \end{aligned}$$

The right hand side goes to 0 as $M \to \infty$ whence

$$\begin{aligned} \langle I(f), I(h)\rangle = \lim_{M\to\infty} \langle I(f_M), I(h_M)\rangle &= \lim_{M\to\infty} \langle \hat{f}_M, \hat{h}_M\rangle \\ &= \langle \hat{f}, \hat{h}\rangle \\ &= \langle f, h\rangle. \end{aligned}$$

□

Now, let $W_t = I(\mathbb{1}_{[0,t]}) = \sum_k g_k \langle \mathbb{1}_{[0,t]}, \psi_k\rangle$ where ψ_k are the Fourier basis. Then,

$$\mathbb{E}(W_t W_s) = \mathbb{E}[I(\mathbb{1}_{[0,t]}) I(\mathbb{1}_{[0,s]})] = \int \mathbb{1}_{[0,\min(s,t)]}\, dx = \min(s,t).$$

Thus, (W_t) is a centered Gaussian process with $\mathrm{Cov}(t,s) = s \wedge t = \min(s,t)$. The following result was given as a homework problem.

Theorem 21. *W_t is a Gaussian process on $\mathbb{R}_+$ with Cov$(t,s) = s \wedge t$ if and only if W_t is a pre-Brownian motion with initial data 0.*

3.2.3 Attempt 3: Levy

Instead of the Fourier basis, we will look at the Haar wavelet basis. Consider the "mother wavelet" $\psi : \mathbb{R} \to \{-1, 1\}$ by

$$\psi(t) = \begin{cases} 1 & 0 < t \leq \frac{1}{2}, \\ -1 & \frac{-1}{2} < t \leq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Define the (n, k)-th Haar function to be $\psi_{n,k}(t) = 2^{n/k}\psi(2^n t - k)$ with $\psi_{0,0}(k) = 1$. Observe that

$$\text{support}(\psi_{n,k}) = \{t : \psi_{n,k}(t) \neq 0\} = \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right].$$

Moreover,

$$\int \psi_{n,k}\psi_{m,\ell}\, dt = \begin{cases} 1 & \text{if } n = m, k = \ell, \\ 0 & \text{otherwise.} \end{cases}$$

One can show that the Haar wavelet basis is in fact an orthonormal basis of $L^2([0, 1])$. Define the Schauder functions $G_{n,k}(t) = \int_0^t \psi_{n,k}(s)\, ds = \langle \mathbb{1}_{[0,t]}, \psi_{n,k}\rangle$.

Theorem 22. *(Levy)*
Let

$$W_t = g_0 t + \sum_{m=1}^{\infty} \sum_{k=0}^{2^m-1} g_{m,k} G_{m,k}$$

then W_t is a Brownian motion. In particular, the series converges uniformly in $\mathcal{C}([0, 1], \|\cdot\|_\infty)$.

Proof. Wiener's isometry 20 implies that W_t is a pre-Brownian motion. Note that

$$\|W_t\|_\infty < \infty \iff \left\| \sum_{m=1}^{\infty} \sum_{k=0}^{2^m-1} g_{m,k} G_{m,k} \right\|_\infty < \infty.$$

Observe that $|G_{m,k}| \leq \frac{1}{2^{m/2}}$. Thus, it is enough to show that

$$\sum_{m=1}^{\infty} \sum_{k=0}^{2^m-1} \frac{|g_{m,k}|}{2^{m/2}} < \infty$$

since the result will follow from the M-test in uniform convergence. Since $\mathbf{g}$ are iid Gaussians,

$$\mathbb{P}(|g_{m,k}| > 2^{m/4}) \leq 2 \cdot \exp\left(-2^{m/2-1}\right)$$

whence

$$\mathbb{P}\left(\max_{1 \leq k \leq 2^m-1} |g_{m,k}| > 2^{m/4}\right) \leq 2^{m+1} \exp\left(-2^{m/2-1}\right).$$

By Borel-Cantelli lemma, there exists m_* so that for $m > m_*$,

$$\sum_k |g_{m,k}| < 2^{m/4} \implies \sum_{m > m_*} \sum \frac{|g_{m,k}|}{2^{m/2}} < \sum \frac{2^{m/4}}{2^{m/2}} < \infty.$$

And so, the series converges uniformly. Continuity of sample paths also follows. $\square$

3.3 Properties of Sample Paths

In this section, we will explore the sample paths of Brownian motion which are not as nice as one would expect. More specifically, it turns out that the sample paths are nowhere differentiable, nowhere monotone, and unbounded. The following result was given as a homework problem.

Proposition 23. *The following are Brownian motions*

1. $-B_t$ *(Symmetry)*
2. $B_t^\lambda = \frac{1}{\lambda} B_{\lambda^2 t}$ *(Scale Invariance)*
3. $B_t^{(s)} = B_t - B_s$ *(Time Translation Invariance)*

Denote $\mathcal{F}_s = \sigma(B_t, t \leq s)$. Observe that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s < t$. Consider the "germ" σ-algebra $\mathcal{F}_{\sigma^+} = \cap_{0<s} \mathcal{F}_s$.

Theorem 24. *(Blumenthal's zero–one law)*
If $A \in \mathcal{F}_{\sigma^+}$ then $\mathbb{P}(A) \in \{0, 1\}$.

Proof. Let $0 < t_1 < \cdots < t_k$ and $g : \mathbb{R}^k \to \mathbb{R}$ be a bounded continuous function. Fix $A \in \mathcal{F}_{\sigma^+}$. Observe that

$$\mathbb{E}[\mathbb{1}_A \cdot g(B_{t_1}, \cdots, B_{t_k})] = \lim_{\varepsilon \to 0} \mathbb{E}[\mathbb{1}_A \cdot g(B_{t_1} - B_\varepsilon, \cdots, B_{t_k} - B_\varepsilon)].$$

However, $(B_{t_1} - B_\varepsilon, \cdots, B_{t_k} - B_\varepsilon) \perp\!\!\!\perp \mathcal{F}_\varepsilon \supseteq \mathcal{F}_{\sigma^+}$. And so,

$$\lim_{\varepsilon \to 0} \mathbb{E}[\mathbb{1}_A \cdot g(B_{t_1} - B_\varepsilon, \cdots, B_{t_k} - B_\varepsilon)] = \mathbb{P}(A) \cdot \mathbb{E}[g(B_{t_1}, \cdots, B_{t_k})].$$

As such, $\mathcal{F}_{\sigma^+} \perp\!\!\!\perp \sigma(B_{t_1}, \cdots, B_{t_k})$ whence $\mathcal{F}_{\sigma^+} \perp\!\!\!\perp \sigma(B_s : s > 0)$. Continuity of Brownian motion implies that $\mathcal{F}_{\sigma^+} \perp\!\!\!\perp \sigma(B_s : s \geq 0) \supseteq \mathcal{F}_{\sigma^+}$. And so, $\mathcal{F}_{\sigma^+} \perp\!\!\!\perp \mathcal{F}_{\sigma^+}$ which means that any event $A \in \mathcal{F}_{\sigma^+}$ satisfies

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2.$$

That is, $\mathbb{P}(A)(1 - \mathbb{P}(A)) = 0$ so $\mathbb{P}(A) \in \{0, 1\}$. □

Theorem 25. *We have the following almost surely:*

1. *For all $\varepsilon > 0$, $\sup_{0<s<\varepsilon} B_s > 0$ and $\inf_{0<s<\varepsilon} B_s < 0$.*
2. *Let $T_a = \inf\{t \geq 0 : B_t = a\}$ then $T_a < \infty$ for all $a \in \mathbb{R}$. In particular,*

$$\limsup_{t \to \infty} B_t = \infty \text{ and } \liminf_{t \to \infty} B_t = -\infty.$$

Proof.

1. Let $\varepsilon_n \downarrow 0$ and $A = \cap_n \{\sup_{0<s<\varepsilon_n} B_s > 0\}$. Since this is a monotone decreasing chain of events, continuity of measure implies that

$$\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}\left(\sup_{0<s<\varepsilon_n} B_s > 0\right) \geq \lim_{n\to\infty} \mathbb{P}\left(B_{\varepsilon_n} > 0\right) \geq \frac{1}{2}.$$

However, $A \in \mathcal{F}_{\sigma^+}$ so Blumenthal's zero–one law 24 implies that $\mathbb{P}(A) = 1$. The other case is similar.

2. Note that

$$1 = \mathbb{P}\left(\sup_{0\leq s\leq 1} B_s > 0\right) = \lim_{\delta\downarrow 0} \mathbb{P}\left(\sup_{0\leq s\leq 1} B_s > \delta\right).$$

On the other hand, scale invariance 23 implies that

$$\mathbb{P}\left(\sup_{0\leq s\leq 1} B_s > \delta\right) = \mathbb{P}\left(\sup_{0\leq s\leq \frac{1}{\delta^2}} \frac{1}{\delta} B_{\delta^2 s} > 1\right) = \mathbb{P}\left(\sup_{0\leq s\leq \frac{1}{\delta^2}} B_s > 1\right).$$

Thus,

$$\lim_{\delta\downarrow 0} \mathbb{P}\left(\sup_{0\leq s\leq 1} B_s > \delta\right) = \lim_{\delta\downarrow 0} \mathbb{P}\left(\sup_{0\leq s\leq \frac{1}{\delta^2}} B_s > 1\right) = \mathbb{P}\left(\sup_{0\leq s} B_s > 1\right).$$

Using scale invariance one more time implies that the right hand side is equal to $\mathbb{P}(\sup_{0\leq s} B_s > M)$ for all $M > 0$. By symmetry, $1 = \mathbb{P}(\inf_{0\leq s} B_s < -M)$. Since Brownian motion is a continuous function, $\limsup_{t\to\infty} B_t = \infty$ and $\liminf_{t\to\infty} B_t = -\infty$. Intermediate Value Theorem then implies that Brownian motion visits all of $\mathbb{R}$. Hence, $T_a < \infty$ for all $a \in \mathbb{R}$. □

Corollary 25.1. *Almost surely, (B_t) is not monotone on any nontrivial interval.*

Proof. By time translation invariance 23/Markov property and symmetry, for any $q \in \mathbb{Q}_+$ and $\varepsilon > 0$, $\sup\limits_{q<s<q+\varepsilon} B_s > B_q$ and same for inf. Because the rationals are dense in $\mathbb{R}$, the result follows. □

Theorem 26. *(Paley–Zygmund/Wiener)*
Almost surely, (B_t) is nowhere differentiable.

Proof. Let

$$\overline{D}f(t) = \lim_{h\to 0^+} \frac{f(t+h)-f(t)}{h},$$
$$\underline{D}f(t) = \lim_{h\to 0^-} \frac{f(t+h)-f(t)}{h}.$$

Then almost surely for all $t > 0$, $\overline{D}B(t) = \infty$ or $\underline{D}B(t) = -\infty$. Indeed, suppose to the contrary that there exists t_0 such that $-\infty < \underline{D}B(t_0) \leq \overline{D}B(t_0) < \infty$. Then

$$\overline{\lim_{h\downarrow 0}} \frac{|B(t_0+h)-B(t_0)|}{h} < \infty$$

so there exists M such that $\sup\limits_{h\in[0,1]} \frac{|B(t_0+h)-B(t_0)|}{h} \leq M$. Suppose that $t \in [\frac{k-1}{2^n}, \frac{k}{2^n}]$ for $n > 2$. By the triangle inequality, for all $1 \leq j \leq 2^n - k$,

$$\begin{aligned}\left|B\left(\frac{k+j}{2^n}\right) - B\left(\frac{k+j-1}{2^n}\right)\right| &\leq \left|B\left(\frac{k+j}{2^n}\right) - B(t_0)\right| + \left|B(t_0) - B\left(\frac{k+j-1}{2^n}\right)\right| \\ &\leq M \cdot \frac{2j+1}{2^n}.\end{aligned}$$

Let $G_{n,k}$ be the event that the above holds for $j = 1, 2, 3$. Then by independence of increments, time translation, and scale invariance, it follows that

$$\begin{aligned}\mathbb{P}(G_{n,k}) &\leq \prod_{j=1}^{3} \mathbb{P}\left(\left|B\left(\frac{k+j}{2^n}\right) - B\left(\frac{k+j-1}{2^n}\right)\right| \leq M \cdot \frac{7}{2^n}\right) \\ &\leq \left[\mathbb{P}(|B(1)|) \leq \frac{7M}{\sqrt{2^n}}\right]^3 \\ &\leq \left(\frac{7M}{\sqrt{2^n}}\right)^3.\end{aligned}$$

The last step is obtained from the fact that

$$\mathbb{P}(|B(1)| \leq x) = 2\int_0^x \frac{e^{-y^2/2}}{\sqrt{2\pi}}\, dy \leq 2\int_0^x \frac{1}{2}\, dy = x.$$

And so,

$$\mathbb{P}\left(\bigcup_{k=1}^{2^n-3} G_{n,k}\right) \leq CM^3 2^n \left(\frac{1}{2^{n/2}}\right)^3 \leq \frac{CM^3}{2^{n/2}}$$

for some constant $C > 0$. Borel Cantelli lemma then implies that $\mathbb{P}(G_{n,k}$ occurs infinitely often$) = 0$. $\square$

Readers can see that the sample paths of Brownian motions are nowhere differentiable, nowhere monotone, and unbounded. Even worse, there does not exist a signed measure so that $B(t) = \nu([0,t]) = \int \mathbb{1}_{[0,t]}\, d\nu$.

3.4 Functional Limit Theory

In this section, we will address the question of in what sense does the simple random walk converge to Brownian motion? Recall that the simple random walk is characterized by $S_n = \sum_{i=1}^n X_i$ where $\mathbb{E}(X_i) = 0$ and $\mathbb{E}(X_i^2) = 1$. In particular, we will show that for any $t \in [0,\infty)$, $\frac{S_{\lfloor nt \rfloor}}{\sqrt{n}} \overset{(D)}{\to} N(0,t)$. However, one first needs a space.

Let $\mathcal{C}([a,b]) = \{f : [a,b] \to \mathbb{R}$ continuous$\}$. Equipped with $\|\cdot\|_\infty$, $(\mathcal{C}([a,b]), \|\cdot\|_\infty)$ is a Banach space. Let

$$\begin{aligned}E_t : \mathcal{C}([0,\infty)) &\to \mathbb{R} \\ f &\mapsto f(t).\end{aligned}$$

The results in this section will be stated without proofs.

Proposition 27. $(\mathcal{C}([0,\infty)),d)$ *is a complete separable metric space where*

$$d(f,g)=\sum_{n=1}^{\infty}\frac{1}{2^n}\left(\max_{0\le x\le n}|f(x)-g(x)|\wedge 1\right).$$

The probability space under consideration in this section is $(\Omega,\mathcal{F})=(\mathcal{C}([0,\infty)),\mathscr{B})$ where $\mathscr{B}$ is the Borel sigma algebra.

Definition 28. *(Law)*
Let $X:(\Omega,\mathcal{F},\mathbb{P})\to(\mathcal{S},\mathscr{S})$*. The law of* X *is* $\mathcal{Q}_X=X_*\mathbb{P}$ *the push forward of* $\mathbb{P}$*. For any* $A\in\mathscr{S}$*,* $\mathcal{Q}_X(A)=\mathbb{P}(X\in A)$*.*

Theorem 29. *If* X *is a* $(\mathcal{C}([0,\infty)),\mathscr{B})$*-valued random variable, its law is uniquely determined by its finite-dimensional distributions.*

Definition 30. *The Wiener measure is the law of Brownian motion denoted by* $\mathcal{Q}_W$*.*

Then the probability space triplet is $(\mathcal{C}([0,\infty)),\mathscr{B},\mathcal{Q}_W)$ and the canonical process is given by $X_t(\omega)=\omega(t)$.

Theorem 31. *(Portmanteau)*
The followings are equivalent

1. $X_n\xrightarrow{D}X$,
2. $F_n(t)\to F(t)$ *at continuous points of* F,
3. $\mathbb{E}f(X_n)\to\mathbb{E}f(X)$ *for continuous bounded* f,
4. $\varphi_{X_n}\to\varphi_X$ *where* φ *is the Fourier transform.*

Suppose that (S,d) is a complete separable metric space and denote $M_1(S)$ to be the collection of probability measures on S.

Definition 32. *We say* $\mu_n\to\mu$ *weakly if* $\int f\,d\mu_n\to\int f\,d\mu$ *for all* $f\in\mathcal{C}_b(S)$ *where* $\mathcal{C}_b(S)$ *denotes the space of continuous bounded functions on* S*.*

Definition 33. *Let* $E\subseteq M_1(S)$ *equipped with the weak convergence topology.*

1. E *is pre-compact if every sequence has a convergence subsequence,*
2. E *is tight if for all* $\varepsilon>0$*, there exists* $K_\varepsilon\subseteq S$ *compact so that* $\mu(K^c)<\varepsilon$ *for all* $\mu\in E$*.*

Theorem 34. *(Prokhorov)*
Let $E\subseteq M_1(S)$*. Then* E *is pre-compact if and only if* E *is tight.*

An interpretation of compactness in infinite dimension is through regularity.

3.5 Donsker's Invariance Principle

Without loss of generality, we will focus on the space $\mathcal{C}([0,1])$ in this section.

Definition 35. *We say $E \subseteq \mathcal{C}([0,1])$*

1. *is uniformly bounded if there exists $M > 0$ so that for all $f \in E$, $\|f\|_\infty \leq M$,*
2. *is equicontinuous if for all $\varepsilon > 0$, there exists $\delta > 0$ such that if $|x-y| < \delta$ then $\sup\limits_{f\in E} |f(x) - f(y)| < \varepsilon$.*

Recall a standard result from real analysis

Theorem 36. *(Arzela-Ascoli)*
Suppose that $E \subseteq \mathcal{C}([0,1])$. Then E is pre-compact if and only if it is uniformly bounded and equicontinuous.

By 2 in 35, it is enough to take $\sup\limits_{f\in E} |f(0)| \leq M$.

Theorem 37. *Suppose that $(\mu_n) \subseteq M_1(\mathcal{C}([0,1]))$. Then (μ_n) is tight if and only if*

1. $\lim\limits_{\lambda\to\infty} \sup\limits_{n\geq 1} \mu_n(|\omega(0)| > \lambda) = 0,$
2. $\lim\limits_{\delta\downarrow 0} \sup\limits_{n\geq 1} \mu_n(\max\limits_{|s-t|<\delta} |\omega(s) - \omega(t)| > \varepsilon) = 0$ *for all $\varepsilon > 0$.*

Proof. The forward direction is left as an exercise. Fix $\eta > 0$ then there exists λ and δ_k so that

$$\sup_{n\geq 1} \mu_n(\underbrace{|\omega(0)| > \lambda}_{A^c}) \leq \frac{\eta}{2},$$

$$\sup_{n\geq 1} \mu_n\left(\underbrace{\max_{|t-s|\leq\delta_k} |\omega(t) - \omega(s)| > \frac{1}{k}}_{B_k^c}\right) \leq \frac{\eta}{2^{k+1}}.$$

Note that $K = \cap_k (A \cap B_k)$ is compact by Arzela-Ascoli theorem. Observe that

$$\begin{aligned}
\mathbb{P}(K) = 1 - \mathbb{P}(K^c) \geq 1 - \sum \mathbb{P}(B_k^c) - \mathbb{P}(A^c) &\geq 1 - \frac{\eta}{2} - \eta \sum \frac{1}{2^{k+1}} \\
&= 1 - \frac{\eta}{2} - \frac{\eta}{2} \\
&= 1 - \eta.
\end{aligned}$$

And so, (μ_n) is tight. ∎

Theorem 38. *Suppose that (X_t^n) is a continuous stochastic process such that*

1. $\sup_n \mathbb{E}(|X_0^n|) < \infty,$

2. There exists $C > 0$ so that $\sup_n \mathbb{E}|X_t^n - X_s^n|^\alpha \leq C|t-s|^{1+\beta}$ for all $0 \leq s < t \leq 1$.

Then their laws (μ_n) are tight.

Now, the question that one might pose is that if (μ_n) is tight and $\mu_{n_k} \to \mu$, $\tilde{\mu}_{n_k} \to \nu$, is $\mu = \nu$?

Theorem 39. *Suppose that $(\mu_n) \subseteq M_1(\mathcal{C}([0,1]))$ is tight. Let X_t^n be the corresponding paths. Suppose that for all k and for all $t_1 < \cdots < t_k$, $\{(X_{t_1}^n, \cdots, X_{t_k}^n)\}$ converges in distribution. Then $\mu_n \to \mu$ weakly such that if $(Y_t) \sim \mu$ then $(X_{t_1}, \cdots, X_{t_k}) \xrightarrow{D} (Y_{t_1}, \cdots, Y_{t_k})$.*

Proof. Since (μ_n) is tight, there exists a subsequence $\mu_{n_k} \to \mu$ weakly. That is, for $\Lambda : \mathcal{C}([0,1]) \to \mathbb{R}$ continuous and bounded,

$$\int \Lambda(\omega)\, d\mu_{n_k}(\omega) \to \int \Lambda(\omega)\, d\mu(\omega).$$

Consider the evaluation map $E_t : \mathcal{C}([0,1]) \to \mathbb{R}$ by $E_t(f) = f(t)$. We could extend E_t to $\mathbb{R}^k$ by letting $E_{t_1,\cdots,t_k}(f) = (f(t_1), \cdots, f(t_k))$. For any $\varphi \in \mathcal{C}_b(\mathbb{R}^k, \mathbb{R})$, consider

$$\begin{aligned} \varphi \circ E_{t_1,\cdots,t_k} : \mathcal{C}([0,1]) &\to \mathbb{R} \\ f &\mapsto \varphi(f(t_1), \cdots, f(t_k)). \end{aligned}$$

Then $\mathbb{E}\varphi(X_{t_1}^{n_k}, \cdots, X_{t_\ell}^{n_k}) \to \mathbb{E}\varphi(\tilde{Y}_{t_1}, \cdots, \tilde{Y}_{t_\ell})$. Thus, every weak limit has the same marginals whence it is the same weak limit. $\square$

Consider the random walk $S_n = \sum_{\ell=1}^k \xi_\ell$ where $\mathbb{E}\xi_\ell = 0$ and $\mathbb{E}\xi_\ell^2 = 1$. For $t \in [0,1]$, let

$$\begin{aligned} Y_t &= S_{[t]} + (t - [t])\xi_{[t]+1}, \\ X_t^n &= \frac{Y_{nt}}{\sqrt{n}}. \end{aligned}$$

Our goal is to show that $(X_t^n) \xrightarrow{D} (B_t)$.

Theorem 40. *(Donsker's Invariance Principle)*
$(X_t^n)_{t\ in[0,1]} \xrightarrow{D} (B_t)_{0 \leq t \leq 1}$.

Proof. **Step 1:** (Finite dimensional distributions) Fix $t_1 < \cdots < t_k$ then $(X_{t_1}, \cdots, X_{t_k}) \xrightarrow{D} (B_{t_1}, \cdots, B_{t_k})$. Recall that if X_n, Y_n, X are random variables in a metric space (S,d) where X_n, Y_n are defined on the same probability space and $X_n \xrightarrow{D} X$ and $d(X_n, Y_n) \xrightarrow{p} 0$, then $Y_n \xrightarrow{D} X$. Now, observe that

$$\left| X_t^n - \frac{1}{\sqrt{n}} S_{[nt]} \right| \leq \frac{1}{\sqrt{n}} \left| \xi_{[nt]+1} \right|$$

whence $d\left(X_t^n, \frac{S_{[nt]}}{\sqrt{n}}\right) \xrightarrow{p} 0$. Without loss of generality, consider $k = 2$ (same argument when $k > 2$ by using induction). Then this means that

$$\left\| (X_{t_1}^n, X_{t_2}^n) - \left(\frac{S_{[nt_1]}}{\sqrt{n}}, \frac{S_{[nt_2]}}{\sqrt{n}} \right) \right\|_2 \xrightarrow{p} 0.$$

From the argument above, it reduces to showing that $\left(\frac{S_{[nt_1]}}{\sqrt{n}}, \frac{S_{[nt_2]}}{\sqrt{n}}\right) \xrightarrow{D} (B_{t_1}, B_{t_2})$. By the continuous mapping theorem, it suffices to show that

$$\left(\frac{S_{[nt_1]}}{\sqrt{n}}, \frac{S_{[nt_2]} - S_{[nt_1]}}{\sqrt{n}}\right) \xrightarrow{D} (B_{t_1}, B_{t_2} - B_{t_1}).$$

However, this follows from the central limit theorem. And so, $(X_{t_1}, \cdots, X_{t_k}) \xrightarrow{D} (B_{t_1}, \cdots, B_{t_k})$.
Step 2: (Tightness) We will assume for simplicity that $\mathbb{E}\xi_i^4 < \infty$ (this can be extended to the case of finite second moment). By theorem 38 38, it suffices to prove the following result

$$\mathbb{E}|X_s^n - X_t^n|^4 \leq C|t-s|^2$$

for $0 \leq s \leq t \leq 1$. Indeed, fix $s, t \in [\frac{i}{n}, \frac{i+1}{n}]$, then $|X_t^n - X_s^n| = \frac{|t-s|}{\sqrt{n}}|\xi_{i+1}|$. Thus,

$$\mathbb{E}|X_t^n - X_s^n|^4 \leq \frac{C}{n^2}|t-s|^4 \leq C|t-s|^2.$$

For general $s < t$, let $t_1 = \frac{k_1}{n} = \frac{\lceil sn \rceil}{n}$ and $t_2 = \frac{k_2}{n} = \frac{\lfloor tn \rfloor}{n}$. It is easy to show that for $a, b, c > 0$,

$$(a+b+c)^4 \leq C(a^4 + b^4 + c^4)$$

where $C > 0$ is a constant (the notation C is abused here). And so,

$$\|X_t - X_s\|^4 \leq C\left(\underbrace{\|X_s - X_{t_1}\|^4}_{(1)} + \underbrace{\|X_{t_1} - X_{t_2}\|^4}_{(2)} + \underbrace{\|X_{t_2} - X_t\|^4}_{(3)}\right).$$

From above, (1)$\leq C|t-s|^2$ and (3)$\leq C|t-s|^2$. On the other hand,

$$\begin{aligned}
(2) = \frac{1}{n^2}\mathbb{E}\left(\sum_{i=nt_1}^{nt_2} \xi_i\right)^4 &= \frac{1}{n^2}\mathbb{E}\left(\sum_{i=1}^{L} \xi_i\right)^4 \qquad\qquad (L = n(t_2 - t_1)) \\
&\leq C \cdot \frac{1}{n^2}(L + L^2) \\
&\leq \frac{C'}{n^2}L^2 = C'(t_2 - t_1)^2.
\end{aligned}$$

And thus,

$$\|X_t - X_s\|^4 \leq C[(t-t_2)^2 + (t_2 - t_1)^2 + (t_1 - s)^2] \leq \tilde{C}|t-s|^2.$$

This gives tightness which completes the proof. □

3.6 Strong Markov Property and Applications

The goal in this section is to study the behaviour of $B_{t+\tau} - B_\tau$ for $t \geq 0$ where τ is random. Denote $\mathcal{F}_t = \sigma(B_s : s \leq t)$ and $\mathcal{F}_\infty = \sigma(B_s : s \geq 0)$.

Definition 41. *An $\mathbb{R}_+$-valued random value T is a stopping time if $\{T \leq t\} \in \mathcal{F}_t$.*

Definition 42. *The T-past σ-algebra is*

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : \text{ for all } t \geq 0, A \cap \{T \leq t\} \in \mathcal{F}_t\}.$$

Observe that $\mathcal{F}_T$ is a σ-algebra, T is $\mathcal{F}_T$-measurable, $\mathbb{1}_{\{T<\infty\}} B_T$ is $\mathcal{F}_T$-measurable, and $\mathbb{1}_{\{s \leq T\}} B_s$ is $\mathcal{F}_T$-measurable.

Theorem 43. *(Strong Markov property)*
Let T be a stopping time with $\mathbb{P}(T < \infty) > 0$. For $t > 0$, let $B_t^T = \mathbb{1}_{\{T<\infty\}}(B_{T+t} - B_T)$. Then under $\mathbb{P}(\cdot|T < \infty)$, (B_t^T) is a Brownian motion starting at 0 and $B_t^T \perp\!\!\!\perp \mathcal{F}_T$.

Proof. Suppose first that $T < \infty$ a.s. From 23, (B_t^T) is a Brownian motion starting at 0. It remains to prove $B_t^T \perp\!\!\!\perp \mathcal{F}_T$.

Fix $A \in \mathcal{F}_T$, $0 \leq t_1 \leq \cdots \leq t_k$, and $F \in \mathcal{C}_b(\mathbb{R}^k)$. Our goal is to show that

$$\mathbb{E}\left[\mathbb{1}_A F\left(B_{t_1}^T, \cdots, B_{t_{k'}}^T\right)\right] = \mathbb{P}(A)\mathbb{E}\left[F\left(B_{t_1}, \cdots, B_{t_{k'}}\right)\right]$$

for all $n \geq 1$ and $t \geq 0$. Denote $[t]_n = \inf\{\frac{k}{2^n} : \frac{k}{2^n} \geq t\}$ then

$$F\left(B_{t_1}^T, \cdots, B_{t_{k'}}^T\right) = \lim_{n\to\infty} F\left(B_{t_1}^{[T]_n}, \cdots, B_{t_{k'}}^{[T]_n}\right) \text{ a.s.}$$

Boundary convergence theorem gives

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_A F\left(B_{t_1}^T, \cdots, B_{t_{k'}}^T\right)\right] \\
&= \lim_{n\to\infty} \mathbb{E}\left[\mathbb{1}_A F\left(B_{t_1}^{[T]_n}, \cdots, B_{t_{k'}}^{[T]_n}\right)\right] \\
&= \lim_{n\to\infty} \sum_{k=0}^{\infty} \mathbb{E}\left[\mathbb{1}\left(\underbrace{A \cap \left\{\frac{k-1}{2^n} < T \leq \frac{k}{2^n}\right\}}_{\in \mathcal{F}_{\frac{k}{2^n}}}\right)\right] F\left(B_{\frac{k}{2^n}+t_1} - B_{\frac{k}{2^n}}, \cdots, B_{\frac{k}{2^n}+t_{k'}} - B_{\frac{k}{2^n}}\right).
\end{aligned}$$

By Markov property, the right hand side is equal to

$$\begin{aligned}
&\lim_{n\to\infty} \sum_{k=0}^{\infty} \mathbb{P}\left(A \cap \left\{\frac{k-1}{2^n} < T \leq \frac{k}{2^n}\right\}\right) \mathbb{E}\left[F\left(B_{t_1}, \cdots, B_{t_{k'}}\right)\right] \\
&= \mathbb{P}(A \cap \{T < \infty\})\mathbb{E}\left[F\left(B_{t_1}, \cdots, B_{t_{k'}}\right)\right]
\end{aligned}$$

which gives independence. □

Theorem 44. *(Reflection Principle)*
For all $t > 0$, let $M_t^* = \max_{s \leq t} B_s$. Then for all $a \geq 0$ and $b < a$,

$$\mathbb{P}(M_t^* \geq a, B_t \leq b) = \mathbb{P}(T_a \leq t, B_t \geq 2a - b).$$

In particular, $(M_t^*) \stackrel{D}{=} (|B_t|)$.

Proof. Recall that $T_a < \infty$ a.s. since Brownian motion hits all points a.s. Then

$$\mathbb{P}(M_t^* \geq a, B_t \leq b) = \mathbb{P}(T_a \leq t, B_t \leq b) = \mathbb{P}(T_a \leq t, B_{t-T_a} \leq b - a).$$

Strong Markov property implies that $(T_a, B^{T_a}) \stackrel{D}{=} (T_a, -B^{T_a})$. And so, the right hand side of the equality above is

$$= \mathbb{P}(T_a \leq t, B_t \geq 2a - b).$$

Also,

$$\begin{aligned}\mathbb{P}(M_t^* \geq a) &= \mathbb{P}(M_t^* \geq a, B_t \geq a) + \mathbb{P}(M_t^* \geq a, B_t \leq a) \\ &= 2\mathbb{P}(B_t \geq a).\end{aligned}$$

□

Theorem 45. *Suppose that (ξ_i) are iid centered with $\mathbb{E}\xi_i^2 = 1$. Then*

$$\lim \mathbb{P}\left(\frac{\max_{k\leq n} S_k}{\sqrt{n}} \geq a\right) = 2\mathbb{P}(B_1 \geq a).$$

Observe that central limit theorem gives $\frac{S_n}{\sqrt{n}} \xrightarrow{D} Z \sim N(0,1)$ which gives $\frac{\max_{k\leq n} S_k}{\sqrt{n}} \xrightarrow{D} |Z|$.

Proof. Note that $F(\omega) = \max_{0\leq s\leq t} \omega(s)$ is a continuous Lipschitz function on $\mathcal{C}([0,1])$. In particular,

$$\begin{aligned}|F(\omega) - F(\omega')| = \left|\max_{s\leq t} \omega(s) - \max_{s\leq t} \omega'(s)\right| &\leq \max_{s\leq t} |\omega(s) - \omega'(s)| \\ &= \|\omega - \omega'\|_\infty .\end{aligned}$$

Using Donsker's invariance principle 40 and Portmanteau's theorem 31 gives the desired result. □

Theorem 46. *$(T_a)_{a\geq 0}$ is a nondecreasing Markov process with transitions*

$$p(a,s) = \frac{a}{\sqrt{2\pi s^3}} exp\left(\frac{-a^2}{2s}\right)$$

and $(T_a) \stackrel{D}{=} \left(\frac{a^2}{B_1^2}\right)$.

Proof. Fix $a \geq b \geq 0$ and let $t \geq 0$. Observe that

$$\begin{aligned}\{T_a - T_b = t\} &= \{B(T_b + s) - B(T_b) < a - b \text{ for all } s < t \text{ and } B(T_b + t) - B(T_b) = a - b\} \\ &\perp\!\!\!\perp \mathcal{F}_{T_b} \text{ and } \{T_d : d \leq b\}.\end{aligned}$$

Thus, (T_a) is a Markov process. By strong Markov property,

$$\mathbb{P}(T_a - T_b \leq t) = \mathbb{P}(T_{a-b} \leq t) = 2\int_{a-b}^{\infty} \frac{1}{\sqrt{2\pi t}} \exp\left(\frac{-x^2}{2t}\right)\, dx.$$

Perform the change of variable $x = \sqrt{\frac{t}{s}}(a - b)$ gives the desired result. □

It is easy to see that Brownian motion is a martingale with respect to $(\mathcal{F}_t)$.

Lemma 47. *$B_t^2 - t$ is a martingale.*

Proof. One has

$$\begin{aligned}\mathbb{E}\left(B_t^2 - t|\mathcal{F}_s\right) &= \mathbb{E}\left[(B_t - B_s)^2 + 2B_tB_s|\mathcal{F}_s\right] - B_s^2 - t \\ &= t - s + 2B_s^2 - B_s^2 - t \\ &= B_s^2 - s.\end{aligned}$$

□

4 Stochastic Integration

4.1 Motivation

In this section, the theory of integration for random functions will be developed. More specifically, we are interested in defining $\int_0^t f(s,\omega)\,dB_s(\omega)$ when f is random. For example, say computing $\int_0^t B_s\,dB_s$. The first thing one might try is using Stieltjes integration.

Let $\mathcal{P}_n$ denote a partition of $[0,T]$ and

$$I_n(f) = \sum_{\substack{[a,b)\in\mathcal{P}_n \\ t^*\in[a,b)}} f(t^*)(B_b - B_a).$$

The integral will likely be $I(f) = \lim I_n(f)$. However, the question is when can one do this?

Definition 48. *(Total variation)*
f is said to have bounded total variation ($f \in BV[0,T]$) if

$$\sup_{\mathcal{P}} \sum_{[a,b)\in\mathcal{P}} |f(b) - f(a)| < \infty.$$

And so, Stieltjes integration will work if $B \in \mathrm{BV}$. This turns out to be false.

Definition 49. *The quadratic variation of B at time t is defined as*

$$\langle B\rangle_t = \lim_{n\to\infty} \sum_{\mathcal{P}_n\subseteq[0,t]} (B_{t_i} - B_{t_{i-1}})^2.$$

Theorem 50. *Let $\mathcal{P}_n$ be an increasing sequence of partitions of $[0,t]$ such that $\sup_{[a,b)\in\mathcal{P}_n} |a - b| \overset{n\to\infty}{\to} 0$. Then*

$$\langle B\rangle_t = \lim_{n\to\infty} \sum (B_{t_i^n} - B_{t_{i-1}^n})^2 = t$$

in L^2.

Proof. It follows from independence of increments that

$$\begin{aligned}
\mathbb{E}\left[\left(\sum (B_{t_i^n} - B_{t_{i-1}^n})^2 - t\right)^2\right] &= \sum \operatorname{Var}\left[\left(B_{t_i^n} - B_{t_{i-1}^n}\right)^2\right] \\
&= 2\sum (t_i^n - t_{i-1}^n)^2 \\
&\leq 2 \sup_i |t_i^n - t_{i-1}^n| \cdot t.
\end{aligned}$$

The right hand side goes to 0 as $n \to \infty$ whence the result follows from squeeze theorem. $\square$

Theorem 51. *$B \notin BV$ a.s.*

Proof. By passing to a subsequence, one can assume a.s. convergence. And so,

$$t = \lim \sum (B_{t_i^n} - B_{t_{i-1}^n})^2 \leq \overline{\lim}\left[\sup_i \left|B_{t_i^n} - B_{t_{i-1}^n}\right| \cdot \sum \left|B_{t_i^n} - B_{t_{i-1}^n}\right|\right].$$

Note that $\sup_i \left|B_{t_i^n} - B_{t_{i-1}^n}\right| \to 0$ s $n \to \infty$ by uniform convergence of Brownian motion. Thus,

$$\lim_{n\to\infty} \sum \left|B_{t_i^n} - B_{t_{i-1}^n}\right| = \infty$$

for otherwise, the inequality would not hold. Hence, $B \notin \mathrm{BV}$. $\square$

Stieltjes integration does not work in this case. In fact, the following example will show that even the simplest approximation breaks.

Let $\mathcal{P}$ be a partition of $[0, T]$. Consider the left and right simple sums of $\int_0^T B_s\, dB_s$

$$\begin{aligned}
\varphi_L(t) &= \sum_i B(t_{i-1}) \mathbb{1}_{[t_{i-1}, t_i)}(t), \\
\varphi_R(t) &= \sum_i B(t_i) \mathbb{1}_{[t_{i-1}, t_i)}(t).
\end{aligned}$$

Then, by independence of increments and completing the square,

$$\begin{aligned}
\mathbb{E}\left(\int_0^T \varphi_L\, dB\right) &= \mathbb{E}\left[\sum_i B(t_{i-1})\left(B_{t_i} - B_{t_{i-1}}\right)\right] = 0 \\
\mathbb{E}\left(\int_0^T \varphi_R\, dB\right) &= \mathbb{E}\left[\sum_i B(t_i)\left(B_{t_i} - B_{t_{i-1}}\right)\right] = \mathbb{E}\left[\sum_i \left(B_{t_i} - B_{t_{i-1}}\right)^2\right] = T.
\end{aligned}$$

The example shows that the issue lies in picking t^*. As such, this leads to different definitions of stochastic integrals. The first one is Ito integral where one picks $t^* = t_{i-1}$ which is the left end point. The second one is Stratonovich integral where one picks $t^* = \frac{t_i + t_{i-1}}{2}$ which is the midpoint. The advantage of Ito integral is that it is a martingale while that of Stratonovich integral is that it plays well with calculus. Ito integral will be explored in the upcoming sections.

4.2 Ito's Integral and Ito's Lemma

Definition 52. *$f : [0, \infty) \times \Omega \to \mathbb{R}$ is progressively measurable if for all $t > 0$, $f : [0, t] \times \Omega \to \mathbb{R}$ is jointly measurable on $([0, t] \times \Omega, \mathscr{B}([0, t]) \times \mathcal{F}_t)$. Alternatively, f is jointly measurable on $([0, \infty) \times \Omega, \mathscr{B} \times \mathcal{F})$ if $f_{[0,t]}$ is $\mathcal{F}_t$-adapted for all t.*

Let

$$V = \left\{ f \text{ progressively measurable : for all } T,\, \mathbb{E}\left(\int_0^T |f(s, \omega)|^2\, ds\right) < \infty \right\}.$$

Theorem 53. *Let $\mathcal{F}$ be the filtration of Brownian motion. For all $f \in V$,*

$$X_f(t, \omega) = \int_0^t f(s, \omega)\, dB_s(\omega)$$

is well-defined and has the following properties

1. *The map $f \mapsto X_f$ is linear,*
2. *X_f is progressively measurable, an $\mathcal{F}$-martingale, and a.s. continuous,*
3. *$X_f^2(t) - \int_0^t |f(s, \omega)|^2\, ds$ is also a martingale.*

Proof. (Sketch) **Step 1:** (Simple functions) Fix $0 = t_0 < t_1 < \cdots < t_n < \infty$. Let $f_j(\omega) \in \mathcal{F}_{t_j} = \sigma(B_s : s \leq t_j)$ be bounded and

$$f(s, \omega) = \begin{cases} f_{j-1}(\omega) & t_{j-1} \leq s < t_j, \\ 0 & t > t_n. \end{cases}$$

For $t \in [t_{k-1}, t_k)$, one has that

$$X_f(t) = \sum_{j=1}^{k-1} f_{j-1}(\omega)(B(t_j) - B(t_{j-1})) + f_{k-1}(B(t) - B(t_{k-1}))$$

with $X_f(t) = X_f(t_n)$ if $t > t_n$. Then 1 and 2 are clear.

For 3, it suffices to show that for $t_k \leq s < t \leq t_{k+1}$,

$$\mathbb{E}\left[X_f^2(t) - X_f^2(s) - \int_s^t f^2\, dx \middle| \mathcal{F}_s \right] = 0.$$

Indeed,

$$\begin{aligned} \mathbb{E}(X_f^2(t) - X_f^2(s)) = \mathbb{E}\left[(X_f^(t) - X_f(s))^2 | \mathcal{F}_s\right] &= \mathbb{E}\left[f_{k-1}^2 (B(t) - B(s))^2 | \mathcal{F}_s\right] \\ &= \mathbb{E}\left[f_{k-1}^2 (t - s) | \mathcal{F}_s\right] \\ &= \mathbb{E}\left[\int_s^t f^2\, dx \middle| \mathcal{F}_s \right]. \end{aligned}$$

To extend, recall Doob's L^2-inequality: If (M_t) is a continuous martingale then for all $T > 0$ and $p > 1$,

$$\mathbb{E}\left[\sup_{s\leq T}|M_s|^p\right] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[|M_T|^p\right].$$

Suppose that $f_n \to f$ in the sense that

$$\mathbb{E}\left[\int_0^T |f_n(s,\omega) - f(s,\omega)|^2\, dx\right] \to 0$$

for all T with X_{f_n} satisfies 1-3. Then,

$$\begin{aligned}\mathbb{E}\left[\sup_{0\leq s\leq T}|X_{f_n}(s) - X_{f_m}(s)|^2\right] = \mathbb{E}\left[\sup_{0\leq s\leq T}|X_{f_n-f_m}(s)|^2\right] &\leq 4\mathbb{E}\left(|X_{f_n-f_m}(T)|^2\right)\\ &= 4\mathbb{E}\left(\int_0^T |f_n - f_m|^2\, dx\right) \overset{m,n\to\infty}{\to} 0.\end{aligned}$$

Step 2: (f is bounded, a.s. continuous, and 0 for $t \geq T_0$). Then denote

$$f_n(s,\omega) = f\left(\frac{j}{n}, \omega\right)$$

for $\frac{j}{n} \leq s \leq \frac{j+1}{n}$. Then $f_n \to f$ pointwise in s ω-a.s. As such, it follows from the bounded convergence theorem that

$$\mathbb{E}\left[\int_0^T |f_n - f|^2\, dx\right] \to 0.$$

Step 3: f is progressively measurable, bounded, and 0 for $t > T_0$ with $f(s,\omega) = 0$ for $s < 0$. Define

$$f_n(s,\omega) = n\int_{s-1/n}^{s} f(x,\omega)\, dx$$

for $s \geq 0$. Observe that f_n is progressively measurable, continuous, bounded, and 0 for $t > T_0'$. Recall the Lebesgue differentiation theorem which states that for $\varphi \in L^2[0,T]$, $\varphi_n = n\int_{s-1/n}^{s}\varphi(x)\, dx$, $\varphi_n \to \varphi$ in $L^2[0,T]$. And so, $f_n \to f$ in the $L^2(dx * \mathbb{P})$ sense.

Step 4: (General) Let

$$f_n(s) = f(s)\mathbb{1}_{[0,n]}(s)$$

then $f_n \to f$. That is, one truncates f when $|f| \geq n$. Then repeat the ideas above will yield the desired result. $\square$

Note that one did not need B_t to be Brownian motion in the proof above. Now, the question is what if one considers a function of Brownian motion, then how will the Ito integral be? To answer this question, the most important formula, namely Ito's formula, will be presented. First, let us explore an example of computing $\int_0^t B_s\, dB_s$.

For fixed $t > 0$, note that $\int_0^t B_s\, dB_s$ in L^2 is the limit of

$$\begin{aligned}
&\sum_{\ell=1}^{2^n} B\left(\frac{\ell-1}{2^n}t\right)\left[B\left(\frac{\ell}{2^n}t\right) - B\left(\frac{\ell-1}{2^n}t\right)\right] \\
&= \frac{1}{2}\sum\left[B\left(\frac{\ell}{2^n}t\right)^2 - B\left(\frac{\ell-1}{2^n}t\right)^2\right] - \frac{1}{2}\sum\left[B\left(\frac{\ell}{2^n}t\right) - B\left(\frac{\ell-1}{2^n}t\right)\right]^2.
\end{aligned}$$

From 50, the right hand side converges in L^2 to

$$\frac{1}{2}[B(t)^2 - B(0)^2] - \frac{t}{2} = \frac{1}{2}\left[B(t)^2 - t\right].$$

If one denotes $f(x) = \frac{x^2}{2}$, then this example shows that

$$\underbrace{\frac{1}{2}B_t^2}_{f(B_t)} = \underbrace{\int_0^t B_s\, dB_s}_{\int_0^t f'(B_s)\, dB_s} + \underbrace{\frac{t}{2}}_{\text{Ito's correction term}}.$$

In particular, Ito integral does not have chain rule in contrast to Stratonovich integral.

Theorem 54. *(Ito's formula)*
Suppose that $f \in \mathcal{C}_b^{1,2}$ (once in time, twice in space), then

$$f(t, B_t) - f(0, B_0) = \int_0^t \partial_s f(s, B_s)\, ds + \int_0^t \partial_x f(s, B_s)\, dB_s + \int_0^t \frac{1}{2}\partial_{xx} f(s, B_s)\, ds.$$

Proof. For $f(t,x) \in \mathcal{C}_b$, it is easy to verify the following result called the QV lemma that

$$\sum_{\substack{\ell \\ 0 \le \ell h \le t}} f(\ell h)[B_{(\ell+1)h} - B_{\ell h}]^2 \xrightarrow{p} \int_0^t f(s, B_s)\, ds$$

as $h \to 0$. Without loss of generality, take $f \in \mathcal{C}_0^\infty$ (smooth and compactly supported). Taylor's theorem implies that

$$\begin{aligned}
f[(\ell+1)h, B((\ell+1)h)] - f(0, B_0) = \partial_s f(\ell h, B_{\ell h})h + \partial_x f(\ell h, B_{\ell h})[B_{(\ell+1)h} - B_{\ell h}] \\
+ \frac{1}{2}\partial_{xx} f(\ell h, B_{\ell h})[B_{(\ell+1)h} - B_{\ell h}]^2 + \text{Error}_\ell.
\end{aligned}$$

The first and second terms in the equality follow from definition of regular integral and Ito integral respectively. The QV lemma implies that the third term converges to $\int_0^t \frac{1}{2}\partial_{xx} f(s, B_s)\, ds$ in probability. On the other hand, Taylor's theorem gives

$$\begin{aligned}
|\text{Error}_\ell| &\le \left|\partial_s \partial_x f(\tau_\ell, \xi_\ell) h (B_{(\ell+1)h} - B_{\ell h})\right| + \frac{1}{3!}\left|\partial_{xxx} f(\tau_\ell, \xi_\ell)(B_{(\ell+1)h} - B_{\ell h})^3\right| \\
&\le C\left[h + (B_{(\ell+1)h} - B_{\ell h})^2\right]\left|B_{(\ell+1)h} - B_{\ell h}\right|
\end{aligned}$$

for some constant C. And so,

$$\sum |\mathrm{Error}_\ell| \le C \underbrace{\sup_{\substack{\ell' \\ 0\le \ell' h\le t}} \left|B_{(\ell'+1)h} - B_{\ell' h}\right|}_{\to 0 \text{ a.s.}} \left(t + \underbrace{\sum \left(B_{(\ell'+1)h} - B_{\ell' h}\right)^2}_{\xrightarrow{L^2} t} \right)$$

Therefore, $\sum |\mathrm{Error}_\ell| \overset{h\to 0}{\to} 0$ in probability. □

The QV lemma shows that $(dB_s)^2 = ds$. Informally, Ito's formula is given along with the following multiplication rule

	dB_t	dt
dB_t	dt	0
dt	0	0

5 Brownian Motion and Partial Differential Equations

In this section, we will explore the relationship between Brownian motions and the theory related to partial differential equations. Throughout this section, denote

$$\mathbb{P}_x(A) = \mathbb{P}(A|B_0 = x),$$
$$\mathbb{E}_x(f(B_t)) = \mathbb{E}(f(B_t)|B_0 = x).$$

The first question that we address is that given $f : \mathbb{R}^d \to \mathbb{R}$, can $\mathbb{E}_x f(B_t)$ be computed?

To answer this question, observe that if f is bounded and $d = 1$, then

$$\mathbb{E}_x f(B_t) = \int f(y) p_t(x, y)\, dy$$

where

$$p_t(x, y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{|x-y|^2}{2t}\right).$$

p_t is called the Heat kernel with $\partial_t p_t(x, y) = \frac{1}{2}\Delta_x p_t(x, y)$. Let $u(t, x) = \mathbb{E}_x f(B_t)$ then

$$\partial_t u(t, x) = \int f(y)\partial_t p_t(x, y)\, dy = \int f(y) \frac{1}{2}\Delta_x p_t(x, y)\, dy = \frac{1}{2}\Delta_x u(t, x).$$

Additionally, $B_t \to B_0$ as $t \to 0$ a.s. whence $\lim_{t\to 0} u(t, x) = f(x)$. As such, $u \in \mathcal{C}^{1,2}$ solves the Heat equation with initial data f

$$\begin{cases} \partial_t u = \frac{1}{2}\Delta u & (t, x) \in (0, \infty) \times \mathbb{R}^d, \\ u(0, x) = f(x). \end{cases}$$

5.1 Dirichlet Problem

Theorem 55. *(Multidimensional Ito's formula)*
Suppose that $f : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ and $f \in \mathcal{C}^{1,2}$ with $\mathbb{E}\int_0^T |\nabla_x f|^2(s, B_s) < \infty$ for all $T > 0$. Then

$$f(t, B_t) - f(0, B_0) = \int_0^t \nabla f(s, B_s) \cdot dB_s + \int_0^t \left(\partial_t + \frac{1}{2}\Delta\right) f(s, B_s)\, ds.$$

The multiplication rule is $dB_t^i dB_t^j = \delta_{ij} dt$, $dB_t^i dt = 0$, and $dtdt = 0$.

If f is nice and $(\partial_t + \frac{1}{2}\Delta)f = 0$, then $f(t, B_t)$ is a martingale if it has moments.

Definition 56. *$D \subseteq \mathbb{R}^d$ is a domain if it is open and connected.*

Definition 57. *Let $D \subseteq \mathbb{R}^d$ be a domain and $u \in \mathcal{C}^2(D)$. We say that u is harmonic on D if $\Delta u = 0$.*

Let D be a bounded domain. The *Dirichlet problem* is finding a function u so that

$$\begin{cases} \Delta u(x) = 0 & x \in D, \\ u(x) = f(x) & x \in \partial D. \end{cases}$$

In particular, if $D \ni y \to x \in \partial D$ then $u(y) \to f(x)$.

Observe that $X_t = u(B_t) - u(x) = \int_0^t \nabla u \cdot dB_s + \int_0^t \frac{1}{2}\Delta u(B_s)\, dB_s$ from Ito's formula. If $x \in D$ and $\tau_D = \inf\{t \geq 0 : B_t \in \partial D\}$, then

$$X_{t\wedge\tau_D} = \int_0^{\tau\wedge\tau_D} \nabla u \cdot dB_s$$

so $\mathbb{E}(X_{\tau_D}) = 0$ whence $u(x) = \mathbb{E}_x u(B_{\tau_D})$. Let's rigorize this idea.

5.2 Solution to Dirichlet Problem I

Proposition 58. *Let u be a harmonic function on a domain D. Let $D' \subseteq D$ be a bounded subdomain. Let $\tau = \tau_{\overline{D'}} = \inf\{t \geq 0 : B_t \in \partial D\}$. Then*

$$u(x) = \mathbb{E}_x u(B_\tau)$$

for all $x \in D'$.

Proof. By Ito's formula,

$$u(B_t) - u(B_0) = \int_0^t \nabla u_s \cdot dB_s + \int_0^t \frac{1}{2}\Delta u(s, B_s)\, ds.$$

Since $\overline{D'}$ is compact, $u, \nabla u \in L^\infty(\overline{D'})$ (bounded) and $\tau < \infty$ p_x-a.s. Hence, $u(B_{t\wedge\tau})$ is a martingale in L^2. By the bounded convergence theorem,

$$u(x) = \mathbb{E}_x u(B_{t\wedge\tau}) \to \mathbb{E}_x u(B_\tau).$$ ◻

Definition 59. *(Mean Value Property)*
Let D be a domain. u satisfies the mean value property on D if for all $x \in D$ and $r > 0$ so that $b_r(x) \subseteq D$, $u(x) = \mathbb{E}u(Y)$ where $Y \sim$ Unf$(\partial b_r(x))$.

Theorem 60. *Let u be bounded measurable on D. u is harmonic if and only u satisfies the mean value property.*

Proof. Suppose u is harmonic then $u \in \mathcal{C}^2(D)$. Because Brownian motion is rotational invariant,

$$T_1 = \inf\{t \geq 0 : \|B_t\| = 1\}$$

so $B_{T_1} | B_0 = 0 \sim \text{Unf}(\partial b_1(0))$. By scaling and translation invariance,

$$T_{x,r} = \inf\{t \geq 0 : \|B_t - x\| = r\}$$

and $B_{T_{x,r}} | B_0 = x \sim \text{Unf}(\partial b_r(x))$. By 58, if $b_r(x) \subseteq D$, $u(x) = \mathbb{E}_x u(B_{T_{x,r}})$ which proves the forward direction.

Conversely, suppose that u is bounded and measurable with mean value property on D. Then $u \in \mathcal{C}^2(D)$ (in fact, $u \in \mathcal{C}^\infty(D)$. One will first perform "molification" of u. Without loss of generality, suppose that $u \equiv 0$ on D^c. Let $\psi : \mathbb{R}^d \to \mathbb{R}_+$ be

$$\psi(x) = \begin{cases} C \cdot \exp\left(\frac{1}{\|x\|^2 - 1}\right) & \|x\|^2 < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then $\psi \in \mathcal{C}_0^\infty$ and support$(\psi) = \overline{b_1(0)}$. C is chosen so that $\int \psi \, dx = 1$. Denote $\psi_\varepsilon(x) = \varepsilon^d \psi(x/\varepsilon)$ which has support $\overline{b_\varepsilon(0)}$ with $\int \psi_\varepsilon \, dx = 1$. The standard molification of u is

$$u_\varepsilon(x) = (\psi_\varepsilon * u)(x) = \int u(y)\psi_\varepsilon(x - y) \, dy.$$

Note that $u_\varepsilon \in \mathcal{C}^\infty(D)$. Fix $x \in D$ and ε so that $b_\varepsilon(x) \subseteq D$. Using polar coordinate gives

$$\begin{aligned} u(x) = \int u(x - y)\psi_\varepsilon(y) \, dy &= C_\varepsilon \int_0^\varepsilon p(r) \int_{\partial b_r(0)} u(x - y) \exp\left(\frac{1}{\frac{r^2}{\varepsilon^2} - 1}\right) dy \, dr \\ &= u(x) \end{aligned}$$

where the last step is obtained from the mean value property. Using Taylor's theorem implies that

$$u(x + y) = u(x) + \sum_i \partial_i u(x) y_i + \frac{1}{2} \sum \partial_i \partial_j u y_i y_j + o(\|y\|^2)$$

for $y \in b_\varepsilon(0)$ where ε is small enough. By symmetry, $y \sim \text{Unf}(\partial b_r(0))$ and

$$\mathbb{E}(Y) = 0, \, \mathbb{E}(Y_i Y_j) = 0, \mathbb{E}Y_i^2 = \frac{r^2}{d}.$$

Thus, if $Y \sim \text{Unf}(\partial b_\varepsilon(0))$,

$$u(x) = \mathbb{E}u(x + Y) = u(x) + \frac{\varepsilon^2}{d} \Delta u + o(\varepsilon^2)$$

implying that $\frac{\varepsilon^2}{d} \Delta u = o(\varepsilon^2)$ whence $\Delta u = 0$. That is, u is harmonic. □

Definition 61. *(classical solution)*
We say that $u : \overline{D} \to \mathbb{R}$ *is a classical solution to the Dirichlet problem if* $u \in \mathcal{C}^2(D) \cap \mathcal{C}(\overline{D})$ *and*

$$\begin{cases} \Delta u = 0 & x \in D, \\ u(x) = g(x) & x \in \partial D. \end{cases}$$

g is called the boundary condition.

Proposition 62. *Let D be a bounded domain.*

1. *Let $g \in \mathcal{C}(\partial D)$ and u be a classical solution to the Dirichlet problem. Then u satisfies $u(x) = \mathbb{E}_x g(B_{\tau_D})$ where τ_D is the exit time of D.*

2. *Let $g \in L^\infty(\partial D)$ (bounded and measurable) then $u(x) = \mathbb{E}_x g(B_{\tau_D})$ is harmonic on D.*

Note that 2 does not mean u solves the Dirichlet problem since there is no guarantee that u is continuous on ∂D.

Proof. 1. Fix $x \in D$ and pick ε_0 so that $b_\varepsilon(x) \subseteq D$ for all $\varepsilon < \varepsilon_0$. Let

$$D_\varepsilon = \{y \in D : d(y, D^c) > \varepsilon\}$$

and D_ε^x be the connected component of D_ε with $x \in D_\varepsilon^x$. Then $\overline{D_\varepsilon^x}$ is compact which implies that u is harmonic on D_ε^x. By Ito's formula, $u(x) = \mathbb{E}_x[u(B_{\tau_{D_\varepsilon^x}})]$. Continuity implies that $\tau_{D_\varepsilon^x} \to \tau_D$ p_x-a.s. By the Bounded Convergence Theorem, $u(x) = \mathbb{E}_x[u(B_{\tau_{D_\varepsilon^x}})] \to \mathbb{E}_x[g(B_{\tau_D})]$.

2. Recall that harmonic is equivalent to having mean value property 60. We will show that u satisfies the mean value property on D. Indeed, fix $x \in D$ and $r > 0$ so that $b_r(x) \subseteq D$. Define

$$T_{x,r} = \inf\{t \geq 0 : \|B_t - x\| = r\}.$$

Starting at x, $T_{x,r} < T_D$ because one must exit the ball before exiting the set. Note that

$$T_D = \underbrace{T_D - T_{x,r}}_{\text{distributed like } T_D | B_0 = B_{T_{x,r}}} + T_{x,r}.$$

Let $\Phi : \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d) \to \mathbb{R}$ with $\gamma_0 \in D$ where $\Phi((\gamma_t)) = g(\gamma_{T_D})$ and $\Phi((B_t)_{t\geq 0}) = \Phi\left((B_{T_{x,r+t}})_{t\geq 0}\right)$ for $B_0 = x$. Then,

$$\begin{aligned} u(x) = \mathbb{E}_x[g(B_{T_D})] = \mathbb{E}_x[\Phi((B_t))] &= \mathbb{E}_x[\Phi((B_{T_{x,r+t}})_t)] \\ &= \mathbb{E}_x[\mathbb{E}_{B_{T_{x,r}}}[\Phi((\tilde{B}_t)_{t\geq 0})]] \\ &= \mathbb{E}_x[u(B_{T_{x,r}})] \end{aligned}$$

where the second last step is from the strong Markov property and the last step is from the definition of u. By symmetry, $B_{T_{x,r}} \sim \mathrm{Unf}(\partial b_r(x))$ so u has mean value property. □

Definition 63. *(Exterior Cone Condition)*
We say that $y \in \partial D$ satisfies the exterior cone condition if there exists a non-empty cone ζ_y with apex y and $r > 0$ so that $\zeta_y \cap b_r(y) \subseteq D^c$.

Theorem 64. *Let D be a bounded domain in $\mathbb{R}^d$. Suppose that for all $y \in \partial D$, y satisfies the exterior cone condition. Then for all $g \in \mathcal{C}(\partial D)$, $u(x) = \mathbb{E}_x g(B_{\tau_D})$ gives a unique classical solution of the Dirichlet problem with boundary condition g.*

Proof. By 62, it suffices to show that for $D \ni x \to y \in \partial D$, $u(x) \to g(y)$. Because ∂D is compact, it follows that g is uniformly continuous on ∂D so for all $\varepsilon > 0$, there exists $\delta > 0$ such that if $|y - y'| < \delta$, $|g(y) - g(y')| < \varepsilon$. Moreover, $\|g\|_\infty \leq M < \infty$ for some constant $M > 0$. Fix ε then one has such a δ. In particular,

$$\begin{aligned}
|u(x) - g(y)| &\leq \mathbb{E}_x\left[|g(B_{\tau_D}) - g(y)|\right] \\
&= \mathbb{E}_x\left[|g(B_{\tau_D}) - g(y)|, \tau_D \leq \eta\right] + \mathbb{E}_x\left[|g(B_{\tau_D}) - g(y)|, \tau_D > \eta\right] \\
&= \underbrace{\mathbb{E}_x\left[|g(B_{\tau_D}) - g(y)|, \tau_D \leq \eta, \sup_{t\leq\eta} d(B_t, x) < \delta/2\right]}_{(1)} + 2M\,\underbrace{\mathbb{P}_x\left[\sup_{t\leq\eta} d(B_t, x) > \delta/2\right]}_{(2)} \\
&\quad + 2M\,\underbrace{\mathbb{P}_x(\tau_D > \eta)}_{(3)}.
\end{aligned}$$

If $|x - y| < \delta/2$, then on $\{\tau_D \leq \eta, \sup_{t\leq\eta} d(B_t, x) < \delta/2\}$, $|B_{\tau_D} - y| < \delta$ so $|(1)| \leq \varepsilon$. On the other hand,

$$(2) = \mathbb{P}_0\left(\sup_{t\leq\eta} |B_t| > \delta/2\right) \overset{\eta\to 0}{\to} 0.$$

Thus, pick $\eta_{\varepsilon,\delta}$ so that $|(2)| \leq \varepsilon$.

Lemma 65. *If y satisfies the exterior cone condition, then for all $\eta > 0$, $\lim_{\substack{x\to y\\ x\in D}} \mathbb{P}_x(\tau_D > \eta) = 0$.*

Proof. Given $v \in \mathbb{R}^d$ with $\|v\| = 1$, consider $\zeta(v, \gamma) = \{z : \frac{z\cdot u}{\|z\|} > 1 - \gamma\}$. The exterior cone condition implies that for any y, there exist v, γ, r with $(y + \zeta(v, \gamma)) \cap b_r(0) \subseteq D^c$.

By Blumenthal's 0-1 law 24,

$$\mathbb{P}_0\left(\tau_{\zeta(v,\gamma/2)\cap b_{r/2}(0)} = 0\right) = 1.$$

Let $\zeta = \zeta(v, \gamma) \cap b_r(0)$ and $\zeta' = \zeta(v, \gamma/2) \cap b_{r/2}(0)$. Then $\zeta'_a := \zeta' \cap b_a(0)^c \uparrow \zeta'$ and $\tau_{\zeta'_a} \downarrow \tau_{\zeta'}$ $\mathbb{P}_0$-a.s. Thus, for all $\beta > 0$, there exists a with

$$\mathbb{P}_0(\tau_{\zeta'_a} \leq \eta) \geq 1 - \beta.$$

However, $y + \zeta \subseteq D$ whence

$$\mathbb{P}_x(\tau \leq \eta) \geq \mathbb{P}_x(\tau_{y+\zeta} \leq \eta) \geq \mathbb{P}(\tau_{y-x+\zeta} \leq \eta) \geq 1 - \beta. \qquad \square$$

Given the lemma, there exists $\delta' < \delta/2$ so that $|x - y| < \delta'$ implies that $(3) < \varepsilon$. As such, if $|x - y| < \delta'$, $(1) + (2) + (3) < 3\varepsilon$. $\square$

5.3 Solution to Dirichlet Problem II

Let $D_{\varepsilon,R} = \{x \in \mathbb{R}^d : \|x\| \in (\varepsilon, R)\}$. Recall that a solution to $\mathrm{Dir}(D, g)$ if exists is $u(x) = \mathbb{E}_x g(B_\tau)$ where $\tau = \inf\{t \geq 0 : B_t \notin D\}$. In this section, restrict $D = b_1(0)$.

Definition 66. *(Poisson Kernel)*
The Poisson kernel is $k : b_1(0) \times \partial b_1(0) \to \mathbb{R}_+$ *with* $k(x, y) = \frac{1-\|x\|^2}{\|x-y\|^d}$.

A tedious calculation can be performed to verify the following result

Lemma 67. *For all* $y \in \partial b_1(0)$, $k_y : x \mapsto k(x, y)$ *is harmonic on* $b_1(0)$.

Definition 68. *Given* $h : (r_1, r_2) \to \mathbb{R}$, *consider* $u : D_{r_1,r_2} \to \mathbb{R}$ *with* $u_h(x) = h(|x|)$. u *is a radial harmonic function if* $u = u_h$ *for some* h *and* u *is harmonic.*

Lemma 69. $u_h : D_{r_1,r_2} \to \mathbb{R}$ *is a radial harmonic function if there exist* a, b *so that*

$$h(r) = \begin{cases} a + b\log r & d = 2, \\ a + br^{2-d} & d \geq 3. \end{cases}$$

Proof. If u_h is radial harmonic, then

$$\begin{aligned} 0 = \Delta u_h &= \Delta h(|x|) \\ &= \nabla \cdot (\nabla h(|x|)) \\ &= \nabla \cdot \left(h'(|x|) \cdot \frac{x}{|x|} \right) \\ &= h''(|x|) \cdot \langle \frac{x}{|x|}, \frac{x}{|x|} \rangle + h'(|x|)\frac{d}{|x|} + h'(|x|)\langle \frac{x}{|x|}, \frac{-x}{|x|^2} \rangle \\ &= h''(|x|) + \frac{d-1}{|x|}h'(|x|). \end{aligned}$$

Solving the differential equation gives the desired result. □

Lemma 70. *For all* $x \in b_1(0)$, $\int k(x, y)\, \sigma_{0,1}(dy) = 1$.

Proof. Let $F(x) = \int k(x, y)\sigma_{0,1}(dy)$. We will show that F is harmonic. Indeed, since k_y is harmonic for all $y \in \partial b_1(0)$,

$$k(x, y) = \int k(z, y)\sigma_{x,r}(dz)$$

for all $r < 1 - |x|$. As such,

$$\begin{aligned} \int F(z)\sigma_{x,r}(dz) &= \iint k(z, y)\sigma_{0,1}(dy)\sigma_{x,r}(dz) \\ &= \iint k(z, y)\sigma_{x,r}(dz)\sigma_{0,1}(dy) \\ &= \int k(x, y)\sigma_{0,1}(dy) \\ &= F(x). \end{aligned}$$

So F satisfies the mean value property implying that it is harmonic. Moreover, F is also radial harmonic. Note that

$$F(0) = \int \frac{1-|0|^2}{|y-0|^2}\sigma(dy) = \int 1\sigma(dy) = 1.$$

Thus $F(x) = 1$. $\square$

Theorem 71. *If $g : \partial b_1(0) \to \mathbb{R}$ is continuous then the unique solution to Dir$(b_1(0), g)$ is*

$$u(x) = \int_{\partial b_1(0)} k(x,y)g(y)\sigma_{0,1}(dy).$$

Proof. u is harmonic by repeating the proof in 70 and replacing F with u. Let $y_0 \in \partial b_1(0)$. If $x \in b_1(0)$, $y \in \partial b_1(0)$, $|x - y_0| < \frac{\delta}{2}$, and $|y - y_0| > \delta$ then $|x - y| > \frac{\delta}{2}$. In particular,

$$k(x,y) = \frac{1-|x|^2}{|x-y|^d} < \left(\frac{2}{\delta}\right)^d \left(1-|x|^2\right).$$

Then for $\delta > 0$,

$$\begin{aligned}\lim_{\substack{x\to y_0\\ x\in b_1(0)}} \int_{|y-y_0|>\delta} k(x,y)\sigma_{0,1}(dy) &\le \lim_{\substack{x\to y_0\\ x\in b_1(0)}} \int_{|y-y_0|>\delta} \left(\frac{2}{\delta}\right)^d \left(1-|x|^2\right)\sigma_{0,1}(dy)\\ &= 0\end{aligned}$$

by the bounded convergence theorem. Since g is continuous and $\partial b_1(0)$ is compact, g is uniformly continuous on $\partial b_1(0)$. As such, for any $\varepsilon > 0$, there exists $\delta > 0$ so that $|y-y_0| < \delta$ implies that $|g(y) - g(y_0)| < \varepsilon$ for all $y, y_0 \in \partial b_1(0)$. Furthermore,

$$M := \sup\{|g(y)| : y \in \partial b_1(0)\} < \infty.$$

Then,

$$\begin{aligned}|u(x) - g(y_0)| &= \left|\int k(x,y)[g(y)-g(y_0)]\sigma_{0,1}(dy)\right|\\ &\le \left|\int_{|y-y_0|\le\delta} k(x,y)[g(y)-g(y_0)]\sigma_{0,1}(dy)\right| + \left|\int_{|y-y_0|>\delta} k(x,y)[g(y)-g(y_0)]\sigma_{0,1}(dy)\right|\\ &\le \int_{|y-y_0|\le\delta} \varepsilon k(x,y)\sigma_{0,1}(dy) + 2M\int_{|y-y_0|>\delta} k(x,y)\sigma_{0,1}(dy)\\ &\le \varepsilon + 2M\int_{|y-y_0|>\delta} k(x,y)\sigma_{0,1}(dy).\end{aligned}$$

As such, for all $\varepsilon > 0$,

$$\lim_{\substack{x\to y_0\\ x\in b_1(0)}} |u(x) - g(y_0)| < \varepsilon.$$ $\square$

As a consequence of this, one has the following results

Corollary 71.1. *Let $T = \inf\{t \geq 0 : B_t \notin b_1(0)\}$. For all $x \in b_1(0)$, a measurable subset A of $\partial b_1(0)$,*

$$\mathbb{E}_x \mathbb{1}_A(B_T) = \mathbb{P}_x(B_T \in A) = \int \mathbb{1}_A(y) k(x,y) \sigma_{0,1}(dy).$$

Proposition 72. *Let $u_a = \inf\{t \geq 0 : |B_t| = a\}$. Let $x \neq 0$ and $0 < \varepsilon < |x| < R$.*

i. $\mathbb{P}_x(u_\varepsilon < u_R) = \begin{cases} \frac{\log R - \log |x|}{\log R - \log \varepsilon} & d = 2, \\ \frac{|x|^{2-d} - R^{2-d}}{\varepsilon^{2-d} - R^{2-d}} & d \geq 3. \end{cases}$

ii. As $\varepsilon \to 0$ *and* $R \to \infty$, $\mathbb{P}_x(u_0 < \infty) = 0$.

iii. As $R \to \infty$,

$$\mathbb{P}_x(u_\varepsilon < \infty) = \begin{cases} 1 & d = 2, \\ \left(\frac{|x|}{\varepsilon}\right)^{2-d} & d \geq 3. \end{cases}$$

6 Stochastic Differential Equations

6.1 Solution Theory

In this section, we are interested in solution theory of stochastic differential equation (SDE) of the form

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t$$

which is a short form for the equation

$$X_t = X_0 + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s. \tag{1}$$

The question one first poses is what does it mean to solve the SDE above? We would first need a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ and an $(\mathcal{F}_t)$-Brownian motion. In general, the terms b and σ are called the drift and volatility respectively. Denote

$$L^\infty_{\text{loc}}(\mathbb{R}^n) = \{f : \text{ for all } K \subseteq \mathbb{R}^n \text{ compact, } f \in L^\infty(K)\}.$$

Definition 73. *Given $b, \sigma \in L^\infty_{loc}$, a solution of the SDE $I(b,\sigma) : dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t$ is*

1. *A filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$,*
2. *An $\mathcal{F}_t$-Brownian motion with $B_0 = 0$,*
3. *An $\mathcal{F}_t$-adapted X_t with continuous sample paths such that (1) holds.*

Definition 74. *Given the $I(b, \sigma)$,*

1. *We say weak existence holds for $I(b,\sigma)$ if for all $x \in \mathbb{R}^d$, there exists a solution $I_x(b,\sigma)$,*

2. *We say weak uniqueness holds for $I(b,\sigma)$ if for all $x \in \mathbb{R}^d$, all solutions of $I_x(b,\sigma)$ are equal in law,*

3. *We say pathwise uniqueness holds if for any solutions X, X' on the same space with respect to the same Brownian motions satisfying $X_0 = X'_0$ a.s. then $(X_t) = (X'_t)$ a.s.,*

4. *We say a strong solution exists if weak existence, pathwise uniqueness, X is adapted to filtration of B.*

A fact we will not prove is that weak existence and weak uniqueness are not sufficient to imply pathwise uniqueness. However, a well-known result by Yamada and Watanabe states that weak existence and pathwise uniqueness imply weak uniqueness. In this section, we will focus on strong solutions.

For the sake of simplicity, we will assume that b and σ are continuous and Lipschitz uniformly in time. That is, there exists $K > 0$ which does not depend on t so that

$$|\sigma(t,x) - \sigma(t,y)| \leq K|x-y|,$$
$$|b(t,x) - b(t,y)| \leq K|x-y|$$

for all t, x, y.

Theorem 75. *Assume that the above assumption holds. Then*

1. *Pathwise uniqueness holds for $I(b,\sigma)$,*

2. *Given a filtered probability space with a Brownian motion, there is a unique strong solution.*

Proof. There are two important results needed to prove this theorem which we state below

i. (Ito's isometry): Let $M_t = \int_0^t a(t,\omega)\, dB_t$ then $M_t^2 - \int_0^t a^2\, ds$ is a martingale and $\mathbb{E}(M_t^2) = \int_0^t \mathbb{E}(a^2)\, ds$.

ii. (Doob's L^2-inequality)

$$\mathbb{P}\left(\sup_{s\leq t}|M_s| > \varepsilon\right) \leq \frac{1}{\varepsilon^2}\cdot \mathbb{E}\left(\int_0^t a^2\, ds\right)$$

or

$$\mathbb{E}\left(\sup_{s\leq t}|M_s|^2\right) \leq C\cdot \mathbb{E}\left(\int_0^t a^2\, ds\right)$$

for some constant $C > 0$.

1. Recall Gronwall's inequality which states that for $a, b \geq 0$, if $g(t) \leq a + b\int_0^t g(s)\, ds$ then $g(t) \leq ae^{bt}$. The strategy that one will employ is that by letting $g(s) = \mathbb{E}\left(\sup_{s\leq t}|X_s - X'_s|^2\right)$, we will aim to show that $g(s) \leq C\int_0^s g(\omega)\, d\omega$ for $0 < s < T$.

Let X, X' be two solutions and fix $T > 0$. Denote

$$\tau = \inf\{t \geq 0 : |X_t| \geq M \text{ or } |X'_t| \geq M\}.$$

Observe that

$$\begin{aligned}
\mathbb{E}\left[(X_{t\wedge\tau} - X'_{t\wedge\tau})^2\right] &= \mathbb{E}\left[\left(\int_0^{t\wedge\tau} \sigma(s, X_s) - \sigma(s, X'_s)\, dB_s + \int_0^{t\wedge\tau} b(s, X_s) - b(s, X'_s)\, ds\right)^2\right] \\
&\leq 2\left[\mathbb{E}\left(\left(\int_0^{t\wedge\tau} \sigma(s, X_s) - \sigma(s, X'_s)\, dB_s\right)^2\right) + \mathbb{E}\left(\left(\int_0^{t\wedge\tau} b(s, X_s) - b(s, X'_s)\, ds\right)\right.\right. \\
&\leq C\left[\mathbb{E}\left(\int_0^{t\wedge\tau} (\sigma(s, X_s) - \sigma(s, X'_s))^2\, ds + T\int_0^{t\wedge\tau} (b(s, X_s) - b(s, X'_s))^2\, ds\right)\right] \\
&\leq CK(1+T)\cdot\mathbb{E}\left(\int_0^{t\wedge\tau} |X_s - X'_s|^2\, ds\right) \\
&= CK(1+T)\int_0^t \mathbb{E}\left(|X_{s\wedge\tau} - X'_{s\wedge\tau}|^2\right)\, ds.
\end{aligned}$$

And so, $g(t) \leq C(T)\int_0^t g(s)\, ds$ for all $t < T$ whence $g(t) \equiv 0$ for $t < T$ by Gronwall's inequality.

2. The idea for the proof is that one will view the solution as a fixed point equation. That is, $X_t = F(X_t)$, $X_t = \underbrace{\int_0^t f(X_t)\, dt}_{F(X_t)}$. Then Picard iteration will be performed with $X_t^0 = x$, $X_t^1 = F(X_t^0)$.

   Let $X_t^0 = x$, $X_t^1 = x + \int_0^t b(s, x)\, ds + \int_0^t \sigma(s, x)\, dB_s$, and

$$X_t^{n+1} = x + \int_0^t b(s, X_s^n)\, ds + \int_0^t \sigma(s, X_s^n)\, dB_s.$$

   Note that $|X_t^n - X_t^{n-1}| \leq C|X_t^{n-1} - X_t^{n-2}|$. Let $g_n(t) = \mathbb{E}\sup_{s\leq t}|X_s^n - X_s^{n-1}|^2$. Doob's inequality gives $g_1(t) \leq C_T$. Then using a similar reasoning as part 1,

$$g_{n+1}(t) = \mathbb{E}\sup_{s\leq t}|X_s^{n+1} - X_s^n|^2 \leq C(1+T)\mathbb{E}\int_0^t |X_\omega^n - X_\omega^{n-1}|^2\, d\omega$$

   which implies that

$$\begin{aligned}
g_{n+1}(t) \leq C(T)\int_0^t g_n(\omega)\, d\omega &\leq C^2(T)\int_0^t\int_0^t g_{n-1}(\omega)\, d\omega \\
&\leq \cdots \\
&\leq C^{n-1}\int_0^t \cdots \int_0^t g_1(\omega)\, d\omega_1 \cdots d\omega_{n-1} \\
&\leq C^{n-1}\cdot C'\cdot\frac{t^{n-1}}{(n-1)!}.
\end{aligned}$$

Therefore, $\sum g_n(t)^{1/2} \le C' e^{C\sqrt{t}} < \infty$ and so

$$\left[\mathbb{E}\sup_{s\le T}\left|\sum_{\ell=1}^{\infty} X_s^\ell - X_s^{\ell-1}\right|^2\right]^{1/2} \le \sum_\ell \left(\mathbb{E}\sup_{s\le T}|X_s^\ell - X_s^{\ell-1}|^2\right)^{1/2} < \infty.$$

As such, $\sum X^\ell - X^{\ell-1}$ converges uniformly a.s. on $\mathcal{C}([0,T])$. $\square$

6.2 Examples

Recall the Markov property

$$\begin{aligned} X_t &= x + \int_0^t b(\omega, X_\omega)\, d\omega + \int_0^t \sigma(\omega, X_\omega)\, dB_\omega \\ &= X_s + \int_s^t b(\omega, X_\omega)\, d\omega + \int_s^t \sigma(\omega, X_\omega)\, dB_\omega \end{aligned}$$

for $s < t$. With $dX_t = \sigma(t, X_t)\, dB_t + b(t, X_t)\, dt$, denote

$$L_t = (\sigma\sigma^T, \nabla^2 \cdot) + (b, \nabla \cdot) = \sum_{i,j} (\sigma\sigma^T)_{ij}\partial_i\partial_j + \sum_j b_j\partial_j.$$

Ito lemma gives

$$f(X_t) - f(X_0) = \int_0^t \langle \nabla f, \sigma\, dB_t\rangle + \int_0^t L_\omega f(X_\omega)\, d\omega.$$

And so, SDEs are equivalent to elliptic parabolic PDEs.

Example 1: Consider the Ornstein–Uhlenbeck process where $\lambda > 0$

$$\begin{cases} dX_t = dB_t - \lambda X_t\, dt \\ X_0 = x. \end{cases}$$

Let $Y_t = e^{\lambda t}X_t$ then

$$\begin{aligned} dY_t = \lambda Y_t\, dt + e^{\lambda t}\, dX_t &= \lambda Y_t\, dt + e^{\lambda t}\, dB_t - \lambda e^{\lambda t} X_t\, dt \\ &= e^{\lambda t}\, dB_t. \end{aligned}$$

So $Y_t = X_0 + \int_0^t e^{\lambda\omega}\, dB_\omega$ and thus,

$$\begin{aligned} X_t = e^{-\lambda t}X_0 + \int_0^t e^{-\lambda(t-s)}\, dB_s &= e^{-\lambda t}X_0 + B\left(\int_0^t e^{-2\lambda(t-s)}\, ds\right) \\ &= e^{-\lambda t}X_0 + B\left(\frac{1-e^{-2\lambda t}}{2\lambda}\right) \\ &\overset{D}{=} e^{-\lambda t}X_0 + \sqrt{\frac{1-e^{-2\lambda t}}{2\lambda}}Z \end{aligned}$$

where $Z \sim N(0,1)$. If $\lambda = \frac{1}{2}$ and $X_0 \sim Z' \perp\!\!\!\perp Z$ then

$$X_t \stackrel{D}{=} e^{-t/2} Z' + \sqrt{1-e^{-t}} Z \stackrel{D}{=} N(0,1).$$

Example 2: For $\sigma, r > 0$, a geometric Brownian motion is given by

$$dX_t = \sigma X_t \, dB_t + r X_t \, dt = X_t(\sigma \, dB_t + r \, dt).$$

One can in fact show that

$$X_t = X_0 \cdot \exp\left(\sigma B_t + \left(r - \frac{\sigma^2}{2}\right) t\right).$$

This process is extremely important in the theory of derivative pricing which is explored in ACTSC 446. Observe that if $r < \frac{\sigma^2}{2}$, then $X_t \downarrow 0$ exponentially fast and $\mathbb{E}(X_t) = x_0 e^{rt} \uparrow \infty$ exponentially fast.

Example 3: Fix $m \geq 0$. The Bessel process is given by

$$dX_t = 2\sqrt{X_t} \, dB_t + m \, dt.$$

Ito lemma gives

$$\begin{aligned} df(X_t) &= f'(X_t) \, dX_t + \frac{1}{2} f''(X_t)(dX_t)^2 \\ &= \left[f'(X_t) m + \frac{1}{2} f''(X_t) \cdot 4 X_t \right] dt + 2 f'(X_t) \sqrt{X_t} \, dB_t \\ &= X_t L f(X_t) \, dt + \cdots \end{aligned}$$

where $Lf = \left(\frac{1}{2}\partial_x^2 f + \frac{m}{x}\partial_x f\right)$ is the Bessel function. If $f(x) = \|x\|^2$ then $f(B_t) = \|B_t\|^2$ and Ito's lemma can be applied. In fact, Ito's lemma still holds when f has polynomial growth. Then

$$\|B_t\|^2 = \|x\|^2 + nt + 2\sum_i \int B_t^i \, dB_t^i.$$

A not so obvious observation is that $\sum_i B_t^i \, dB_t^i \stackrel{D}{=} \|B_t\| \, d\beta_t$ where β_t is a Brownian motion. And so, consider instead the entry-wise process $dY_t = n \, dt + \sqrt{Y_t} \, dB_t$. The idea is that for $n \geq 2$, $\mathbb{P}_x(T_0 < \infty) = 0$ and

$$\|B_t\| = \sum_i \frac{B^i \, dB_i}{\|B\|} \implies \beta_s = \int \frac{\sum_i B_s^i \, dB_s^i}{\|B_j\|} \, ds.$$

Hence,

$$(dB_s)^2 = \left(\frac{\sum_i B_s^i \, dB_s^i}{\|B_j\|}\right)^2 = \frac{\sum_i B_s^2}{\|B_s\|^2} \, dt = dt.$$

Theorem 76. *(Dubins-Schwarz)*
M_t is a continuous martingale with $(dM_t)^2 = dt$ if and only if M_t is a Brownian motion.

In fact, M_t is a continuous martingale and $M_t^2 - t$ is a continuous martingale if and only if M_t is a Brownian motion.

Let

$$f(x) = \begin{cases} x^{-m/2} & m \neq 2, \\ \log(x) & m = 2. \end{cases}$$

Then $df(X_t) = 2X_t^{\frac{1-m}{2}}\, dB_t$. One can look at $X_{t\wedge u_\varepsilon \wedge u_R}$ where u_ε and u_R are defined in 72.

Definition 77. *Let T be a stopping time and $(X_t)_{0\leq t\leq T}$ be adapted. It is called a local martingale if there exist $T_1 < T_2 < \cdots < T_n \to T$ a.s. so that $(X_{t\wedge T_n})$ is a martingale.*

Theorem 78. *Let $D \subseteq \mathbb{R}^d$ be a domain, τ_D be the exit time of D, f be a harmonic function on D. Then $f(B_{t\wedge\tau_D})$ is a local martingale.*

Proof. Let (K_n) be an exhaustion of D by compact. That means $K_n \subseteq K_{n+1} \subseteq \cdots$, $\cup K_n = D$, and K_n are compact. Denote $\tau_n = \tau_{K_n}$ then $f(B_{t\wedge\tau_n})$ is a martingale. And so, $\tau_n \to \tau_D$ a.s. with $f(B_{t\wedge\tau_n}) \to f(B_{t\wedge\tau_D})$. $\square$

With f as follows

$$f(x) = \begin{cases} \|x\|^{-d} & d \neq 2, \\ \log\|x\| & d = 2. \end{cases}$$

it follows that $\mathbb{E}\log\|B_t\|^2 \to \infty$ when $d = 2$ and $\mathbb{E}\|B_t\|^{-d} \to 0$ for $d \geq 3$.

7 General Theory of Markov Process

7.1 Transition Semigroup

In this section, we will focus on semigroups. Consider $Q_t : \mathcal{B}(\mathbb{R}) \to \mathcal{B}(\mathbb{R})$ where $\mathcal{B}(\mathbb{R})$ denotes the space of bounded real-valued functions and $Q_t f(x) = \mathbb{E}_x[f(B_t)]$. Then

$$\begin{aligned} Q_{t+s}f(x) = \mathbb{E}_x f(B_{t+s}) &= \mathbb{E}_x\left[\mathbb{E}_{B_s}[f(B_t)]\right] \\ &= \mathbb{E}_x Q_t f(B_s) \\ &= Q_s Q_t f(x). \end{aligned}$$

One can show that Q_t satisfies

1. $Q_{t+s} = Q_t Q_s$ (semigroup),
2. $Q_0 = I$,
3. If $0 \leq f \leq 1$, then $0 \leq Q_t f \leq 1 + Q_t 1 = 1$,
4. $Q_t : \mathcal{C}_0(\mathbb{R}) \to \mathcal{C}_0(\mathbb{R})$. (Feller)

The key point is that for $f \in \mathcal{C}_0^\infty$,

$$Q_t f(x) = \mathbb{E}_x f(B_t) = f(x) + \int_0^t \underbrace{\mathbb{E}_x \left[\frac{\Delta}{2} f(B_s)\right]}_{Q_s(\frac{\Delta}{2} f)} ds.$$

Thus, $\frac{d}{dt} Q_t f = Q_t \left(\frac{\Delta}{2} f\right)$ whence

$$\begin{aligned} Q_t (\Delta f) = \int \Delta f(y) \cdot p_t(x-y)\, dy &= \int f(y) \Delta p_t(x-y)\, dy \\ &= \Delta Q_t f. \end{aligned}$$

In particular, this means that one has a system of equations

$$\begin{cases} Q_t = \frac{\Delta}{2} Q_t \\ Q_0 = I \end{cases} \implies Q_t = e^{t\Delta/2}.$$

As an example, consider the simple random walk on a graph G. Denote Q to be the transition matrix (normalized adjacency) and the graph Laplacian $\mathcal{L} = Q - I$. Let $Y_t = X_{N(t)}$ where $N(t) \sim \text{Poisson}(t)$. It follows from the Taylor series expansion of e^x that

$$\begin{aligned} Q_t f(x) = \mathbb{E}_x[f(Y_t)] &= \sum_{k=0}^\infty \mathbb{E}_x[f(X_k)] \cdot \mathbb{P}(N(t) = k) \\ &= \sum_{k=0}^\infty (Q_k f)_x \cdot e^{-t} \cdot \frac{t^k}{k!} \\ &= e^{-t} e^{tQ} f(x) = e^{t\mathcal{L}} f(x). \end{aligned}$$

Hence, $Q_t = \mathcal{L} Q_t$. Let us formalize all of this using measure theory.

Definition 79. *(Transition kernel)*
Let $(E, \mathcal{E})$ be a measurable space. A Markov transition kernel from E to E is a map $Q : E \times \mathcal{E} \to [0,1]$ so that

1. *For all $x \in E$, $A \mapsto Q(x, A)$ is a probability measure,*
2. *For all $A \in \mathcal{E}$, $x \mapsto Q(x, A)$ is $\mathcal{E}$-measurable.*

Example 1: $(E, \mathcal{E}) = \left([n], 2^{[n]}\right)$ where $[n] = \{1, \cdots, n\}$ and Q is a transition matrix.

Example 2: $(\mathbb{R}, \mathscr{B})$ and $Q(x, A) = \int_A p_t(x-y)\, dy$ for all t.

Observe that $Qf(x) = \int f(y) Q(x, dy)$ is bounded measurable if f is bounded measurable.

Definition 80. *A collection $(Q_t)_{t \geq 0}$ of transition kernels is called a Markov (transition) semigroup if*

1. *For all $x \in E$, $Q_0(x,dy) = \delta_x$ (i.e. $Q_0 = I$),*

2. *For all s,t, $Q_{t+s}(x,A) = \int_E Q_t(x,dy)Q_s(y,A)$ (i.e. $Q_{t+s} = Q_tQ_s$),*

3. *For all $A \in \mathcal{E}$, $(t,x) \mapsto Q_t(x,A)$ is $\mathscr{B}([0,\infty)] \bigotimes \mathcal{E}$-measurable.*

If one has a transition semigroup on a Polish space (complete separable metric space), then there exists a Markov process (X_t) so that $Q_t(x,A) = \mathbb{P}(X_t \in A | X_0 = x)$. Denote

$$\mathcal{C}_0(E) = \left\{ f \text{ continuous: with } K_n \uparrow E \text{ a compact exhaustion, } \sup_{x \in E \setminus K_n} |f(x)| \to 0 \right\}.$$

In turns out that $(\mathcal{C}_0(E), \|\cdot\|_\infty)$ is a Banach space.

Definition 81. *(Feller semigroup)*
A semigroup (Q_t) is Feller if

1. *$Q_t : \mathcal{C}_0(E) \to \mathcal{C}_0(E)$,*

2. *For $f \in \mathcal{C}_0(E)$, $\|Q_tf - f\|_\infty \to 0$ as $t \to 0$ (strong continuity).*

Definition 82. *(Infinitestimal generator)*
Let

$$\mathcal{D}(\mathcal{L}) = \{f \in \mathcal{C}_0(E) : \mathcal{L}f = g \in \mathcal{C}_0(E)\}$$

where

$$\mathcal{L}f(x) = \lim_{t \to 0} \frac{Q_tf - f}{t}(x) = \lim_{t \to 0} \frac{Q_tf - Q_0f}{t}(x)$$

is called the infinitestimal generator.

Observe that $\mathcal{L}$ is linear and $\mathcal{D}(\mathcal{L})$ is a linear subspace of $\mathcal{C}_0(E)$.

Lemma 83. *If $f \in \mathcal{D}(\mathcal{L})$, then $Q_sf \in \mathcal{D}(\mathcal{L})$ and $\mathcal{L}Q_sf = Q_s(\mathcal{L}f)$ for all s.*

Proof. Note that for $f \in \mathcal{D}(\mathcal{L})$

$$\frac{Q_tQ_sf - Q_sf}{t} = Q_s\left(\frac{Q_tf - f}{t}\right) \overset{t\to 0}{\to} Q_s(\mathcal{L}f)$$ □

Proposition 84. *For $f \in \mathcal{D}(\mathcal{L})$ and $t > 0$,*

$$Q_tf = f + \int_0^t Q_s(\mathcal{L}f)\,ds = f + \int_0^t \mathcal{L}(Q_sf)\,ds.$$

Proof. This follows from the fact that $\frac{1}{h}(Q_{t+h}f - f) \to Q_t(\mathcal{L}f)$. □

7.2 Feller Process

Definition 85. *A Markov process is Feller if its transition semigroup is Feller.*

A fact that we will accept is that every Feller process has the strong Markov property.

An example of a Feller process is the Levy process.

Definition 86. (X_t) *is called a Levy process if*

1. $X_0 = 0$ *a.s.,*
2. X_t *has independent increments,*
3. $X_t - X_s \stackrel{D}{=} X_{t-s}$ *for all* $t > s$*,*
4. X_t *is continuous in probability. That is,*

$$X_{t+h} - X_t \to 0$$

   in probability as $h \to 0$*.*

In fact, if a Levy process is Gaussian then it is a Brownian motion. If it is Poisson$(\lambda(t-s))$ then it is a Poisson process.

Definition 87. *The* λ*-resolvent of* Q_t *is* $R_\lambda f(x) = \int_0^\infty e^{-\lambda t} Q_t f(x)\, dt$*.*

Proposition 88. R_λ *satisfies*

1. $\|R_\lambda f\|_\infty \le \frac{1}{\lambda} \|f\|_\infty$*,*
2. *For* $0 \le f \le 1$*,* $0 \le R_\lambda f \le 1$*,*
3. $R_\lambda - R_\mu + (\lambda - \mu) R_\lambda R_\mu = 0$ *(resolvent identity).*

Proof. We will leave the proofs of 1 and 2 as exercises. Using Fubini's theorem and Markov property, it follows that

$$\begin{aligned}
R_\lambda R_\mu f(x) = \int_0^\infty e^{-\lambda t} Q_t \left(\int_0^\infty e^{-\mu s} Q_s f\, ds \right) dt &= \int_t^\infty e^{-\lambda t} \int_0^\infty e^{-\mu s} Q_{t+s} f\, ds\, dt \\
&= \int_0^\infty e^{-(\lambda-\mu)t} \int_t^\infty e^{-\mu r} Q_r f\, dr\, dt \\
&= \int_0^\infty \int_0^r e^{-(\lambda-\mu)t}\, dt \cdot e^{-\mu r} Q_r f\, dr \\
&= \int_0^\infty \frac{e^{-\mu r} - e^{-\lambda r}}{\lambda - \mu} Q_r f\, dr \\
&= \frac{R_\mu - R_\lambda}{\lambda - \mu} f(x).
\end{aligned}$$

□

Proposition 89. *Fix* $\lambda > 0$*. Then*

1. *$\mathcal{R} = Image(R_\lambda)$ does not depend on λ,*

2. *$\overline{\mathcal{R}} = \mathcal{C}_0(E)$.*

Proof. 1. For $\mu \neq \lambda$, 88 gives

$$R_\mu f = R_\lambda \left(f - (\lambda - \mu) R_\mu f\right)$$

and the result follows.

2. For $f \in \mathcal{C}_0(E)$, observe that

$$\lambda R_\lambda f = \int_0^\infty \lambda e^{-\lambda t} Q_t f \, dt = \int_0^\infty e^{-t} Q_{\frac{t}{\lambda}} f \, dt \overset{\lambda \to \infty}{\to} f$$

because Q is Feller. Hence, $f \in \overline{\mathcal{R}}$ and the result follows. $\square$

Proposition 90. *$\mathcal{D}(\mathcal{L}) = \mathcal{R}$. Furthermore, for all $\lambda > 0$, $R_\lambda g = f$ if and only if $(\lambda - \mathcal{L})^{-1} f = g$.*

Proof. Suppose that $g \in \mathcal{C}_0(E)$. Note that

$$\begin{aligned} \frac{1}{\varepsilon}\left[Q_\varepsilon R_\lambda g - R_\lambda g\right] &= \frac{1}{\varepsilon}\left[\left(1 - e^{-\lambda \varepsilon}\right) \int_0^\infty e^{-\lambda t} Q_{t+\varepsilon} g \, dt - \int_0^\varepsilon e^{-\lambda t} Q_t g \, dt\right] \\ &\to \lambda R_\lambda g - g. \end{aligned}$$

And so, $R_\lambda g \in \mathcal{D}(\mathcal{L})$ and $(\lambda - \mathcal{L}) R_\lambda g = g$ whence $\mathcal{R} \subseteq \mathcal{D}(\mathcal{L})$ and the forward direction holds.

Conversely, fix $f \in \mathcal{D}(\mathcal{L})$ then

$$Q_t f = f + \int_0^t Q_s \mathcal{L} f \, ds$$

and

$$\begin{aligned} R_\lambda f = \int_0^\infty e^{-\lambda t} Q_t f \, dt &= \frac{1}{\lambda} f + \int_0^\infty e^{-\lambda t} \int_0^t Q_s \mathcal{L} f \, ds \, dt \\ &= \frac{f}{\lambda} + \int_0^\infty \frac{e^{-\lambda s}}{\lambda} Q_s \mathcal{L} f \, ds. \end{aligned}$$

Therefore, $f = R_\lambda[(\lambda - \mathcal{L}) f]$ so that $\mathcal{D}(\mathcal{L}) \subseteq \mathcal{R}$. Hence, the backward direction holds. $\square$

Corollary 90.1. *$(\lambda - \mathcal{L})^{-1}$ is a bounded operator of $\mathcal{C}_0(E)$.*

Proof. $\left\|(\lambda - \mathcal{L})^{-1} f\right\|_\infty = \left\|R_\lambda f\right\|_\infty \leq \frac{1}{\lambda} \left\|f\right\|_\infty$. $\square$

Theorem 91. *A Feller semigroup is uniquely determined by its generator.*

Proof. There are two approaches to prove the theorem. the first approach uses the observation that $R_\lambda g$ uniquely solves the equation $(\lambda - \mathcal{L})h = g$. Thus, one can use Laplace inversion to continue. However, we will use the second approach here because it is constructive.

Let $f \in \mathcal{C}_0(E)$ and $t > 0$. We aim to show that

$$\lim_{n\to\infty} \left(I - \frac{t}{n}\mathcal{L}\right)^{-n} f = Q_t f.$$

Indeed, since $(\lambda - \mathcal{L})^{-1} f = R_\lambda f$, it follows that $\left(I - \frac{\mathcal{L}}{\lambda}\right)^{-1} f = \lambda R_\lambda f$. Thus,

$$\left(I - \frac{\mathcal{L}}{\lambda}\right)^{-n} f = \lambda^n R_\lambda^n f = \cdots = \int_0^\infty \underbrace{\frac{\lambda^n s^{n-1}}{(n-1)!} e^{-\lambda s}}_{\text{density of Gamma}(n,\lambda)} \quad Q_s f\, ds.$$

In particular, because $\text{Gamma}(n, \lambda) \overset{D}{=} \sum_{i=1}^n \exp(\lambda) \overset{D}{=} \sum_{i=1}^n \frac{\exp(1)}{\lambda}$, it follows that

$$\left(I - \frac{t}{n}\mathcal{L}\right)^{-n} f = \mathbb{E}\left[Q_{\frac{\sum_{i=1}^n \tau_i}{n}\cdot t} f\right]$$

where $\tau_i \overset{\text{iid}}{\sim} \exp(1)$. Observe that

$$\left\| \left(I - \frac{t}{n}\mathcal{L}\right)^{-n} f - Q_t f \right\|_\infty = \left\| \mathbb{E}\left[Q_{T_t} f - Q_t f\right] \right\|_\infty \le \mathbb{E}\left[\|\mathcal{L}f\|_\infty \cdot |T_t - t|\right] \to 0$$

by the law of large number for all $f \in \mathcal{D}(\mathcal{L})$ since $\overline{\mathcal{D}(\mathcal{L})} = \mathcal{C}_0(E)$. $\square$

Theorem 92. *Let X be a Feller process with generator $\mathcal{L}$. For all $f \in \mathcal{D}(\mathcal{L})$, $M_t^f = f(X_t) - \int_0^t \mathcal{L}f(X_s)\, ds$ is a $\mathbb{P}^x$-martingale for all $x \in E$.*

Proof. Observe that

$$\begin{aligned}
\mathbb{E}\left[M_t^f | \mathcal{F}_s\right] &= \mathbb{E}\left[f(X_t) - \int_s^t \mathcal{L}f_\omega\, d\omega - \int_0^s \mathcal{L}f_\omega\, d\omega | \mathcal{F}_s\right] \\
&= \mathbb{E}\left[f(X_t) - \int_s^t \mathcal{L}f_\omega\, d\omega | \mathcal{F}_s\right] - \int_0^s \mathcal{L}f_\omega\, d\omega \\
&= \mathbb{E}_{X_s}\left[f(X_{t-s}) - \int_0^{t-s} \mathcal{L}f_\omega\, d\omega\right] - \int_0^s \mathcal{L}f_\omega\, d\omega \\
&= Q_{t-s} f(X_s) - \int_0^{t-s} Q_\omega \mathcal{L}f(X_\omega)\, d\omega - \int_0^s \mathcal{L}f_\omega\, d\omega.
\end{aligned}$$

However, $Q_s \mathcal{L} f = \partial_s Q_s f$ so the right hand side is equal to $f(X_s) - \int_0^s \mathcal{L}f(X_\omega)\, d\omega$. $\square$

Theorem 93. *Let $X, \mathcal{L}, Q_t$ as before. Suppose $\tilde{\mathbb{P}}$ is a probability measure on Ω with $\tilde{\mathbb{P}}(X_0 = x) = 1$ and $M_t^f = f(X_t) - \int_0^t \mathcal{L}f(X_s)\, ds$ is a $\tilde{\mathbb{P}}$-martingale for all $f \in \mathcal{D}(\mathcal{L})$. Then $\tilde{\mathbb{P}} = \mathbb{P}^X$.*

Proof. Given $g \in \mathcal{C}_0(E)$, $(\lambda - \mathcal{L})f = \lambda g$ implies that

$$\tilde{\mathbb{E}}\left[f(X_t) - f(X_s) - \int_s^t \mathcal{L}f_\omega \, d\omega | \mathcal{F}_s\right] = 0.$$

Therefore,

$$\tilde{\mathbb{E}}\left[\int_s^\infty \lambda e^{-\lambda t} f(X_t)\, dt - \underbrace{\int_s^\infty \lambda e^{-\lambda t} \int_s^t \mathcal{L}f_\omega \, d\omega \, dt}_{(1)} | \mathcal{F}_s\right] = e^{-\lambda s} f(X_s).$$

In particular,

$$\begin{aligned}
(1) &= \int_s^\infty \mathcal{L}f_\omega \int_\omega^\infty \lambda e^{-\lambda t}\, dt\, d\omega = \int_s^\infty e^{-\lambda\omega}\, \mathcal{L}f_\omega \, d\omega, \\
\text{LHS} &= \tilde{\mathbb{E}}\left[\int_s^\infty e^{-\lambda t}(\lambda - \mathcal{L})f(X_t)\, dt | \mathcal{F}_s\right] - \tilde{\mathbb{E}}\left[\int_s^\infty \lambda e^{-\lambda t} g | \mathcal{F}_s\right].
\end{aligned}$$

Thus,

$$\int_0^\infty \lambda e^{-\lambda t} \tilde{\mathbb{E}}[g(X_{t+s})]\, dt = \tilde{\mathbb{E}} f(X_s)$$

and so $\tilde{\mathbb{E}}g(X_t) = \mathbb{E}g(X_t)$. That is, X_t has the same law under $\tilde{\mathbb{P}}$ and $\mathbb{P}^X$. By usual tricks, finite-dimensional distributions are the same as well. □

8 The Martingale Problem

In the previous section, we have given an analytic approach to diffusion problem. In particular, we have shown that Feller processes are equivalent to semigroups which are determined by infinitesimal generators. In this section, a more probabilistic approach will be explored.

Suppose that $X \sim I(a, b)$ which is given by

$$\begin{cases} dX_t = \sqrt{a}(t, \omega)\, dB_t + b(t, \omega)\, dt \\ X_0 = x \end{cases}$$

then for all $f \in \mathcal{C}_0^\infty(\mathbb{R}^d)$,

$$M_t^f = f(X_t) - f(X_0) - \int_0^t L_\omega f(\omega, X_\omega)\, d\omega$$

is a martingale with

$$L_t = \frac{1}{2}\sum_{ij} a^{ij}(t, \omega)\partial_i\partial_j + \sum_j b^j(t, \omega)\partial_j.$$

Stroock and Varadhan have shown that Ito processes are equivalent to generators via martingales (duality). Given

$$a : \mathbb{R}_+ \times \mathbb{R}^d \to S_d,$$
$$b : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$$

where S_d is the space of symmetric and positive definite $d \times d$ matrices, let

$$L_t = \frac{1}{2} \sum_{ij} a^{ij}(t, \underline{x}) \partial_i \partial_j + \sum_j b^j(t, \underline{x}) \partial_j.$$

An interesting question that is posed is given L_t as above, is there a family $(\mathbb{P}_{s,x})$ on $\mathcal{M}_1(\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d), \mathscr{B})$ so that

1. $\mathbb{P}(X_0 = x) = 1$,
2. For all $f \in \mathcal{C}_0^\infty$, M^f is a $\mathbb{P}_x$-martingale.

Definition 94. *Given $(s, x) \in \mathbb{R}_+ \times \mathbb{R}^d$, a measure $\mathbb{P} \in \mathcal{M}_1(\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d), \mathscr{B})$ is called a solution to the martingale problem for L_t if*

1. $\mathbb{P}(X_0 = x) = 1$,
2. $f(X_t) - \int_s^t L_\omega f(\omega, X_\omega)\, d\omega$ *is a $\mathbb{P}$-martingale for all $f \in \mathcal{C}_0^\infty$.*

Now, the issue that arises is the well-posedness of this problem which consists of existence, uniqueness, and continuous dependence of the solution.

Theorem 95. *The following are equivalent*

1. *X_t is a.s. continuous and*

$$Y_t^i = X_t^i - X_0^i - \int_0^t b^i(\omega, X_\omega)\, d\omega,$$
$$Z_t^{ij} = Y_t^i Y_t^j - \int_0^t a^{ij}(\omega, X_\omega)\, d\omega$$

   are both $\mathbb{P}$-martingales.

2. *For all $\lambda \in \mathbb{R}^d$,*

$$Z_\lambda(t) = exp\left[\langle \lambda, Y_t \rangle - \frac{1}{2} \int_0^t \langle \lambda, a\lambda \rangle \, d\omega\right]$$

   is a martingale.

3. *Same as above but for $\lambda \mapsto i\lambda$ (Fourier transform).*
4. *For all $f \in \mathcal{C}_l^\infty(\mathbb{R}^\lceil)$,*

$$f(X_t) - f(X_0) - \int_0^t L_\omega f(\omega, X_\omega)\, d\omega$$

   is a martingale.

5. For all $f \in \mathcal{C}_b^{1,2}$,

$$f(t, X_t) - f(0, X_0) - \int_0^t (\partial_\omega + L_\omega) f(\omega, X_\omega)\, d\omega$$

is a martingale.

This theorem generalizes Levy characterization of Brownian motion. That is, X_t is a Brownian motion if and only if X_t^i, $X_t^i X_t^j - \delta_{ij} t$ are continuous martingales. Moreover, it gives a new notion of solution in which strong solution implies martingale solution which implies weak solution. In particular, 95 part 5 implies the same but with $f \in \mathcal{C}_b^{1,2}$ and derivatives of at most exponential growth.

Definition 96. *A family $(\mathbb{P}_{s,x})$ is Feller continuous if for all $(s_n, x_n) \to (s, x)$, $\mathbb{P}_{s_n, x_n} \to \mathbb{P}_{s,x}$ weakly.*

Theorem 97. *Let a, b be as above. Suppose that there exists $\sigma : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ so that $\sigma\sigma^T = a$ and there exists C so that*

$$\|\sigma(t,x) - \sigma(t,y)\| + |b(t,x) - b(t,y)| \le C \|x - y\|$$

for all t, x, y. Then the martingale problem for $L = \frac{1}{2}(a, \nabla^2) + (b, \nabla)$ is well-posed. That is, there is a unique solution $(\mathbb{P}_{s,x})$ and this family is Feller continuous.

This gives a new perspective on convergence. We can now provide another proof for Donsker's invariance principle.

Recall the simple random walk on $\mathbb{Z}^d$. (X_k) is a Markov chain on $\mathbb{Z}^d \subseteq \mathbb{R}^d$ so that

$$p(x,y) = \begin{cases} \frac{1}{2d} & \text{if } \|x - y\|_1 = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Consider $X^\delta = \sqrt{\delta} X$ which is a simple random walk on $\sqrt{\delta}\mathbb{Z}^d$ and $Y_t^\delta = X^\delta_{\frac{t}{\delta}}$ continuously interpolated in between. Let $\hat{\mathcal{L}}_\delta = \mathcal{P}_\delta - I$ where $\mathcal{P}_\delta$ is the transition matrix for X^δ. Doob's decomposition gives

$$f(X_n^\delta) - f(X_0^\delta) - \sum_{\ell=0}^{n} \hat{\mathcal{L}}_\delta f(X_\ell)$$

which is a martingale. For $f \in \mathcal{C}_0^\infty(\mathbb{R}^d)$ and $\mathcal{L}_\delta = \frac{1}{\delta}\hat{\mathcal{L}}_\delta$,

$$\mathcal{L}_\delta f(x) = \frac{1}{2d\delta} \sum_i \left[f(x + \sqrt{\delta} e_i) + f(x - \sqrt{\delta} e_i) - 2f(x) \right] \to \frac{1}{2} \Delta f(x).$$

Suppose $Y_t^\delta \to Y_t$ weakly then

$$\begin{aligned} f(Y_t^\delta) - f(Y_0^\delta) - \int_0^t \mathcal{L}_\delta f(Y_\omega^\delta)\, d\omega &\approx f(X^\delta_{\frac{t}{\delta}}) - f(X_0^\delta) - \sum_{\ell=0}^{\frac{t}{\delta}} \hat{\mathcal{L}}_\delta f(X_\ell^\delta) \\ &\to f(Y_t) - f(Y_0) - \int_0^t \frac{1}{2} \Delta f(Y_\omega)\, d\omega \end{aligned}$$

is a martingale for all $f \in \mathcal{C}_0^\infty$. And so, Y_t is a Brownian motion.

9 Scaling Limits of Markov Chains

Suppose that (X_n^h) is a family of time-homogeneous Markov chain in $\mathbb{R}^d$ with $\mathbb{P}^h(x, dy)$. The question that we tackle in this section is when does $\mathbb{P}_x^h \to \mathbb{P}_x$ diffusion weakly?

Let Y_t^h be the interpolant (time step size is h). The key quantities are

$$\begin{aligned}
\mathcal{L}_h f(x) &= \frac{1}{h}\int [f(y) - f(x)]\mathbb{P}_h(x, dy) \\
a_h(x) &= \frac{1}{h}\int_{\|y-x\|\le 1} (y-x)(y-x)^T \mathbb{P}_h(x, dy) \\
b_h(x) &= \frac{1}{h}\int_{\|y-x\|\le 1} (y-x)\mathbb{P}_h(x, dy).
\end{aligned}$$

Morally,

$$\begin{aligned}
\mathbb{E}_x\left[(Y_h - Y_0)^2\right] &\approx a(x)\cdot h, \\
\mathbb{E}_x\left(Y_h - Y_0\right) &\approx b(x)\cdot h.
\end{aligned}$$

We will make the following assumptions

1. There exist a, b continuous so that $a_h \to a$, $b_h \to b$ uniformly on compacts.
2. $\Delta_h^\varepsilon(x) = \frac{1}{h}\mathbb{P}_h(x, b_\varepsilon(x)^c) \to 0$ uniformly on compacts.

Theorem 98. *Assumptions 1 and 2 hold if and only if for all $f \in \mathcal{C}_0^\infty$, $\mathcal{L}_h f \to Lf$ uniformly with*

$$L = \frac{1}{2}\langle a, \nabla^2\cdot\rangle + \langle b, \nabla\cdot\rangle = \frac{1}{2}\sum_{ij} a^{ij}(x)\partial_i\partial_j + \sum_j b_j(x)\partial_j.$$

Proof. Taylor's theorem implies that for all $\varepsilon > 0$ and $y \in b_\varepsilon(x)$,

$$f(y) = f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2}\langle \nabla^2 f(x)(y-x), y-x\rangle + r_f(y)\left\|y-x\right\|^2$$

such that

$$\lim_{\varepsilon\to 0} R(\varepsilon) = \lim_{\varepsilon\to 0}\sup_{y\in b_\varepsilon(x)} r_f(y) = 0.$$

Observe that

$$\begin{aligned}
\mathcal{L}_h f(x) &= \frac{1}{h}\left[\int_{\|y-x\|\le 1}[f(y)-f(x)]\cdot\mathbb{P}_h(x,dy) + \int_{\|y-x\|>1}[f(y)-f(x)]\cdot\mathbb{P}_h(x,dy)\right] \\
&= L_h f(x) + C\cdot\left[\frac{1}{h}\left(\int_{b_1(x)} r_f(y)\cdot\|y-x\|^2\,\mathbb{P}_h(x,dy) + \int_{b_1(x)^c}[f(y)-f(x)]\cdot\mathbb{P}_h(x,dy)\right)\right] \\
&= L_h f(x)
\end{aligned}$$

where

$$L_h = \frac{1}{2}\langle a_h, \nabla^2\rangle + \langle b_h, \nabla\rangle.$$

Note that $L_h f \to L_f$ by assumption 1. On the other hand,

$$\int_{b_1(x)^c} [f(y) - f(x)] \cdot \mathbb{P}_h(x, dy) \le 2\|f\|_\infty \Delta_h^1(x).$$

The right hand side goes to 0 by assumption 2. Furthermore, using Holder's inequality gives

$$\begin{aligned}\frac{1}{h}\int_{b_1(x)} r_f(y)\cdot\|y-x\|^2\,\mathbb{P}_h(x,dy) &= \frac{1}{h}\left(\int_{b_\varepsilon(x)} r_f(y)\cdot\|y-x\|^2\,\mathbb{P}_h(x,dy) + \int_{b_\varepsilon(x)^c\cap b_1(x)} r_f(y)\cdot\|y-x\|^2\,\mathbb{P}_h(x,\right.\\ &\le R(\varepsilon)\cdot\|a_h(x)\| + R(1)\cdot 1\cdot\Delta_h^\varepsilon(x).\end{aligned}$$

By assumptions 1 and 2, the right hand side goes to 0 as $\varepsilon \to 0$. The forward direction then follows.

Conversely, in order to show that assumption 2 holds, it suffices to show that if $x_h \to x_0$ then for all $\varepsilon > 0$, $\Delta_h^\varepsilon(x_h) \to 0$. Let $\chi_\varepsilon = \chi_{b_{\varepsilon/4}(x_0)}$ be the cut off function. In particular, $\chi_\varepsilon \in \mathcal{C}_0^\infty$, $0 \le \chi_\varepsilon \le 1$, $\chi_\varepsilon \equiv 1$ on $b_{\varepsilon/4}(x_0)$, and $\chi_\varepsilon \equiv 0$ on $b_{\varepsilon/2}(x_0)^c$. In particular, $x \in b_{\varepsilon/4}(x_0)$. Now,

$$-\mathcal{L}_h\chi_\varepsilon(x) = \frac{1}{h}\int 1 - \chi_\varepsilon(y)\mathbb{P}_h(x, dy) \ge \Delta_h^{\varepsilon/2}(x).$$

However, $-\mathcal{L}_h\chi_\varepsilon \to -L_h\chi_\varepsilon(x) = 0$ uniformly on compact. Thus, assumption 2 holds.

Denote $\chi_R = \chi_{b_R(0)}$ so that $\chi_R \equiv 1$ on $b_R(0)$ and $\chi_R \equiv 0$ on $b_{2R}(0)^c$. Let $f = x_i$ or x_ix_j and $f_R = f\cdot\chi_R$. Then

$$\begin{aligned}&\mathcal{L}_h f_R(x) \to Lf_R(x) \text{ uniformly on compact or}\\ &\mathcal{L}_h f(x) \to Lf(x) \text{ by using assumption and taking } R \text{ large.}\end{aligned}$$

As such,

$$\mathcal{L}_h f_R(x) = \frac{1}{h}\int f_R(y) - f_R(x)\mathbb{P}_h(x,dy) = \begin{cases} a_h^{ij} \\ b_h^i \end{cases} + \frac{1}{h}\int_{\|x-y\|\ge 1} f_R(y) - f_R(x)\mathbb{P}_h(x,dy)$$

for large enough R and $|\text{second term}| \le 2\|f_R\|_\infty \cdot \Delta_h^1(x) \to 0$ as $h \to 0$. $\square$

Theorem 99. *Suppose assumptions 1 and 2 hold. Then $(\mathbb{P}_h(x,\cdot))$ is precompact and any limit point solves the martingale problem for a, b. In particular, if there exists a unique solution to the martingale problem, $\mathbb{P}_x^h \to \mathbb{P}_x$ weakly (uniformly on compact).*

The key points to prove this theorem are as follows. Fix $k > 0$ and let φ be the cut off for $[-1, 1]$ (vanish on $[-2, 2]^c$),

$$\mathbb{P}_h^k(x, dy) = \varphi\left(\frac{x}{k}\right)\mathbb{P}_h(x, dy) + \left(1 - \varphi\left(\frac{x}{k}\right)\right)\delta_x(dy).$$

One can show that $(\mathbb{P}_h^k(x,\cdot))$ is tight. Consider

$$\begin{aligned}
Z_f^h(nh) &= f(X_n^{h,k}) - f(X_0^{h,k}) - \sum_{j=0}^{h-1} h\mathcal{L}_h f(X_k^{h,k}) \qquad \text{(is a martingale for all } f \in \mathcal{C}_0^\infty)\\
&= f(Y_{nh}^{h,k}) - f(Y_0^{h,k}) - \int_0^{nh} \mathcal{L}_h f(Y_\omega^{h,k})\, d\omega\\
&\to f(Y_t^k) - f(Y_0^k) - \int_0^t (\varphi L) f(Y_\omega^k)\, d\omega.
\end{aligned}$$

Y^k solves the martingale problem for $\varphi(x/k)L = L^k$ with law Q^k. As such, Q^k agrees with $\mathbb{P}_x$ on $\mathcal{F}_{\tau_k}$. However,

$$\left\{\sup_{0\le s\le T} |\omega(s)| \ge k\right\} \in \mathcal{F}_{\tau_k}$$

and it is closed in $\mathcal{C}\left(\mathbb{R}_+, \mathbb{R}^d\right)$. Thus,

$$\overline{\lim}\, \mathbb{P}_{h,x}^k \left(\sup_{0\le s\le T} |\omega(s)| \ge k\right) \le \mathbb{P}_x \left(\sup_{s\le T} |\omega(s)| \ge k\right) \overset{k\to\infty}{\to} 0.$$

On $\mathcal{F}_T$, $\mathbb{P}_h(x,\cdot) - \mathbb{P}_h^k(x,\cdot) \to 0$ and so $\mathbb{P}_h(x,\cdot) \to \mathbb{P}_x$.

10 Girsanov's Theorem and Skorokhod's Embedding Theorem

10.1 Girsanov's Theorem

Let $\mathbb{P}_x, \mathbb{Q}_x$ be the laws of x_t, y_t and consider the equations

$$\begin{aligned}
dx_t &= \sigma(t,x_t)\, dB_t + b(t,x_t)\, dt\\
dy_t &= \sigma(t,y_t)\, dB_t + c(t,y_t)\, dt
\end{aligned}$$

with $x(0) = y(0) = x$. Suppose that $\sigma > 0$ and $\frac{c(t,x)-b(t,x)}{\sigma(t,x)} = e(t,x)$ is bounded.

Theorem 100. *For all x, t, $\mathbb{P}_x, \mathbb{Q}_x$ are mutually absolutely continuous on $\mathcal{F}_t$ and the Radon-Nykodym derivative is given by*

$$\left.\frac{d\mathbb{Q}_x}{d\mathbb{P}_x}\right|_{\mathcal{F}_t} = R_t(\omega) = exp\left[\int_0^t e(s,x_s)\, d\beta_s - \frac{1}{2}\int_0^t e^2(s,x_s)\, ds\right]$$

where

$$d\beta = \frac{1}{\sigma(t,x_t)}\left[dx_t - b(t,x_t)\, dt\right]$$

is a $\mathbb{Q}$-Brownian motion.

Consider an example where $dx_t = dW_t$ and $dy = dB_t + c(t, y_t)\,dt$ with $e = c$. Then $\mathbb{P}_x$ is the Wiener measure, $\mathbb{Q}_x|_{\mathcal{F}_t} = R_t \mathbb{P}_x|_{\mathcal{F}_t}$, and

$$R_t = \exp\left[\int_0^t c(s, W_s)\,dW_s - \frac{1}{2}\int_0^t c(s, W_s)^2\,ds\right].$$

Note that in the two equations, one needs to have the same volatility to avoid being mutually singular. We will omit the proof of this theorem. Readers can find the proof for this theorem in any probability textbook.

10.2 Skorokhod's Embedding Theorem

Let $X \sim \mathrm{Unf}(\{\pm 1\})$. Recall that with T being the exit time of $[-1, 1]$, $B_T \overset{D}{=} X$, $\mathbb{E}(T) < \infty$ whence $\mathbb{E}(B_T) = 0$ and $\mathbb{E}(B_T^2) = \mathbb{E}(T) = \mathbb{E}(X^2)$.

Definition 101. *A martingale (X_n) is a binary splitting if for all $x_0, \cdots, x_n$, $A(x_0, \cdots, x_n) = \{X_0 = x_0, \cdots, X_n = x_n\}$ with $\mathbb{P}(A(x_0, \cdots, x_n)) > 0$ then $X_{n+1}|_{A(x_0,\cdots,x_n)}$ takes at most two values.*

Lemma 102. *Let $X \in L^2$, then there exists binary splittings $X_n \to X$ a.s. in L^2.*

Proof. Let $G_0 = \{\varnothing, \Omega\}$, $x_0 = \mathbb{E}(X)$, and

$$\xi_0 = \begin{cases} 1 & X \geq \mathbb{E}(X), \\ -1 & X < \mathbb{E}(X). \end{cases}$$

Denote $G_n = \sigma(\xi_0, \cdots, \xi_{n-1})$, $x_n = \mathbb{E}(X|G_n)$ and

$$\xi_n = \begin{cases} 1 & X \geq X_n, \\ -1 & X < X_n. \end{cases}$$

Observe that G_n is generated by a partition $\mathcal{P}_n$ of Ω with $|\mathcal{P}_n| = 2^n$. If $A \in \mathcal{P}_n$ then A has the form of $A(x_0, \cdots, x_n)$ so there exist $A_1, A_2 \in \mathcal{P}_{n+1}$ such that $A = A_1 \cup A_2$. And so, one has a binary splitting. Thus, $X_n \to X_\infty = \mathbb{E}(X|G_\infty)$ a.s. in L^2 where $G_\infty = \sigma(\cup G_n)$. We claim that

$$\lim_{n\to\infty} \xi_n(X - X_{n-1}) = |X - X_\infty|.$$

Indeed, if $X = X_\infty$ then we are done. Otherwise, if $X < X_\infty$ then there exists N so that $X < X_n$ for $n \geq N$. And so, $\xi_n = -1$ for all $n \geq N$ whence $\xi_n(X - X_{n-1}) = |X - X_{n-1}|$. This also holds for $X > X_\infty$. In particular,

$$\mathbb{E}\left[\xi_n(X - X_{n-1})\right] = \mathbb{E}\left[\xi_n \mathbb{E}[X - X_{n-1}|G_{n-1}]\right] = 0.$$

Therefore, $\xi_n(X - X_{n-1}) \to 0$ in L^1 which implies that $X_\infty = X$. □

Theorem 103. *Let X be in L^2 with $\mathbb{E}(X) = 0$. Then there exists a stopping time T such that $B(T) \overset{D}{=} X$ and $\mathbb{E}(T) = \mathbb{E}(X^2)$.*

Proof. The key points are as follows. If $X \in [a,b]$ where $a < 0 < b$ then let T be the exit time of $[a,b]$ so that $B_T \overset{D}{=} X$. Let X_n be the binary splitting of X and consider $T_0 \leq T_1 \leq T_2 \leq \cdots$ with $T_n \uparrow T$ a stopping time. Then

$$\begin{aligned}&B(T_n) \to B(T),\\ &\mathbb{E}(T) = \lim \mathbb{E}(T_n) = \lim \mathbb{E}(X_n) = \mathbb{E}(X).\end{aligned}$$

□

11 Invariant Measure, Ergodicity, Dirichlet Forms, and Reversibility

11.1 Invariant Measure

Let $\mathcal{P}_t$ be a Feller semigroup.

Definition 104.

1. *A (σ-finite) measure μ is an invariant measure for $\mathcal{P}_t$ if $\int \mathcal{P}_t f\, d\mu = \int f\, d\mu$ for bounded measurable f. That is, the law of $X_t | X_0 \sim \mu = \mu$ informally.*
2. *If μ is an invariant measure $\mu \in \mathcal{M}_1$, then this implies the existence of stationary distribution.*

Example:

1. Brownian motion $\implies$ $\mu =$ Lebesgue
2. Ornstein–Uhlenbeck process $\implies$ $\mu = N(0, I)$
3. Brownian motion on S' $\implies$ $\mu = \mathrm{Unf}(S')$

The question is can we characterize this?

Example: Let $U = \mathcal{C}_0$, $V = \{\text{finite measures}\}$ with $\langle f, \mu\rangle = \int f\, d\mu$. Recall that the adjoint of $A : U \to U$ is A^* such that

$$\langle u, A^* v\rangle = \langle Au, v\rangle.$$

In particular, note that $\int \mathcal{P}_t f\, d\mu = \int f\, d\mu$ so

$$\lim_{t\to 0} \frac{\mathcal{P}_t f - f}{t}\, d\mu = \int \mathcal{L} f\, d\mu = 0$$

for $f \in \mathcal{D}(\mathcal{L})$. That is, μ solves the adjoint equation $\langle f, \mathcal{L}^*\mu\rangle = \langle \mathcal{L} f, \mu\rangle = 0$ so $\mathcal{L}^*\mu = 0$.

Suppose that $\mu(dx) = \varphi(x)\, dx$ which has density with respect to the Lebesgue measure.

Thus, for all $f \in \mathcal{C}_0^\infty$, $\langle Lf, \varphi\rangle_{L^2(dx)} = 0$ and $L = \frac{1}{2}a(x)\partial_x^2 + b(x)\partial_x$. If φ is regular enough, then

$$0 = \int \left[\frac{1}{2}a(x)\partial_x^2 f(x) + b(x)\partial_x f\right]\varphi(x)\,dx = \int f\left[\frac{1}{2}\partial_x^2(a\varphi) - \partial_x(b\varphi)\right]dx$$

using integration by parts. The adjoint equation is given by

$$\frac{1}{2}\partial_x^2(a\varphi) - \partial_x(b\varphi) = 0.$$

Suppose that μ exists such that $L^2(\mu)$ makes sense. We can define the adjoint by $\mathcal{P}_t^* : L^2 \to L^2$ by

$$\langle f, \mathcal{P}_t^* g\rangle = \langle \mathcal{P}_t f, g\rangle.$$

Observe that $\mathcal{P}_t^*$ is a Markov semigroup.

11.2 Reversibility

Given X_t and $T > 0$, the time reversal Markov process is $(\hat{X}_t)_{t\in[0,T]}$ where $\hat{X}_t = X_{T-t}$. Formally,

$$\begin{aligned}\int \mathcal{P}_t f g\,d\mu = \int \mathbb{E}_x f(X_t)g(X_0)\,d\mu = \int \mathcal{P}_{T-t}\left[\mathcal{P}_t f g\right](x)\,d\mu \\ = \int \mathbb{E}_x\left[\mathcal{P}_t f(X_{T-t})g(X_{T-t})\right]d\mu \\ = \int \mathbb{E}_x\left[f(X_T)g(X_{T-t})\right]d\mu \\ = \int \mathbb{E}_x\left[f(\hat{X}_0)g(\hat{X}_t)\right]d\mu.\end{aligned}$$

An interesting question is how does one make sense of reversible processes?

Theorem 105. *The following are equivalent*

1. $(X_t)_{0\le t\le T} \overset{D}{=} (\hat{X}_t)_{0\le t\le T}$ *(time reversible) if* $X_0 \sim \mu$.
2. $\mathcal{P}_t$ *satisfies detailed balance condition* $\mu(dx)p_t(x,dy) = \mu(dy)p_t(y,dx)$.
3. $\mathcal{P}_t$ *is self-adjoint* $\mathcal{P}_t = \mathcal{P}_t^*$.
4. $\mathcal{L}$ *is self-adjoint such that* $\mathcal{D}(\mathcal{L}) = \mathcal{D}(\mathcal{L}^*)$ *and* $\mathcal{L}f = \mathcal{L}^* f$ *for all* $f \in \mathcal{D}(\mathcal{L})$.

Proof. (2) $\implies$ (1): this follows easily.

(1) $\implies$ (2):

$$\begin{aligned}\mathbb{E}_\mu f(X_0, X_t) = \mathbb{E}_\mu f(\hat{X}_0, \hat{X}_t) &= \mathbb{E}_\mu f(X_T, X_{T-t}) \\ &= \mathbb{E}_\mu f(X_t, X_0)\end{aligned}$$

and $\mu(dx)p_t(x,dy)$ is a symmetric measure.

(2) $\iff$ (3) For all $f,g \in \mathcal{C}_0^\infty$,

$$\begin{aligned}\langle f, \mathcal{P}_t g\rangle = \iint f(x)g(y)p_t(x,dy)\,d\mu(x) &= \iint f(x)g(y)p_t(y,dx)\,d\mu(y)\\ &= \langle \mathcal{P}_t f, g\rangle.\end{aligned}$$

(3) $\implies$ (4)

$$\begin{aligned}\langle \mathcal{L}f, g\rangle = \lim\langle \frac{\mathcal{P}_t - I}{t} f, g\rangle &= \lim\langle f, \frac{\mathcal{P}_t - I}{t} g\rangle\\ &= \langle f, \mathcal{L}g\rangle.\end{aligned}$$

(4) $\implies$ (3):

$$\begin{aligned}\mu(s) = \langle \mathcal{P}_s f, \mathcal{P}_{t-s} g\rangle &\implies \frac{d}{ds}\mu(s) = \langle \mathcal{L}\mathcal{P}_s f, \mathcal{P}_{t-s}g\rangle - \langle \mathcal{P}_t f, \mathcal{L}\mathcal{P}_{t-s} g\rangle = 0\\ &\implies \langle \mathcal{P}_t f, g\rangle = \mu(t) = \mu(0) = \langle f, \mathcal{P}_t g\rangle.\end{aligned}$$

□

11.3 Ergodicity

Definition 106. *A stationary distribution μ for a Markov process is called Ergodic if for all $B \in \mathcal{B}$, if $\mathcal{P}_t \mathbb{1}_B = \mathbb{1}_B$ for all $t \geq 0$ then $\mu(B) \in \{0,1\}$.*

In particular, invariant measures are the convex hull of Ergodic measures.

Define $A_t f = \frac{1}{t}\int_0^t f(X_s)\,ds$. One can check that

1. $\mathbb{E}A_t f = \int f\,d\mu$.
2. $\mathrm{Var}(A_t f) = \frac{2}{t}\int_0^t (1-\frac{r}{t})\mathrm{Cov}_{\mathcal{P}_\mu}\left[f(X_0), f(X_r)\right]\,dr$.
3. Let $f_0 = f - \int f\,d\mu$ then $\int_0^\infty \langle f_0, \mathcal{P}_s f_0\rangle\,ds < \infty$. And so,

$$\lim_{t\to\infty} t\mathrm{Var}(A_t f) = \sigma_f^2 = 2\int_0^\infty \langle f_0, \mathcal{P}_s f_0\rangle\,ds = 2\int_0^\infty \mathrm{Cov}_{\mathcal{P}_\mu}[f(X_0), f(X_s)]\,ds.$$

The question now is in what sense does $\mathcal{P}_t f \to \int f\,d\mu$?

Definition 107. *(L^2 Ergodicity)*
Var($\mathcal{P}_t f$) $\to 0$ and $\int \left(\mathcal{P}_t f - \int f\,d\mu\right)^2\,d\mu \to 0$ for all $f \in L^2(\mu)$.

11.4 Dirichlet Forms

Definition 108. *The Dirichlet form is given by $\mathcal{E}(f,g) = \langle f, -\mathcal{L}g\rangle$.*

Note that $1 \in \mathcal{D}(\mathcal{L})$ with $\mathcal{L}1 = 0$. In particular, $\mathcal{L}f = 0$ if and only if f is constant.

Assume that $-\mathcal{L}$ has pure point spectrum

$$0 \le \lambda_0(-\mathcal{L}) \le \lambda_1(-\mathcal{L}) \le \cdots$$

Courant–Fischer theorem has shown that

$$\lambda_1(-\mathcal{L}) = \min_{\substack{\|f\|_2=1 \\ \langle f,1\rangle\mu=0}} \langle f, -\mathcal{L}f\rangle.$$

Lemma 109. *(Poincare inequality)*
$Var(f) \le \frac{1}{\lambda_1}\mathcal{E}(f,f)$ and $Var(\mathcal{P}_t f) \le e^{-2\lambda_1 t}\, Var(f)$.

Proof. Without loss of generality, assume that $\int f\, d\mu = 0$ anmd define $\psi(t) = \mathbb{E}\left(\mathcal{P}_t f\right)^2$. Observe that

$$\psi'(t) = 2\langle \mathcal{L}\mathcal{P}_t f, \mathcal{P}_t f\rangle \le -2\lambda_1(-\mathcal{L})\underbrace{\langle \mathcal{P}_t f, \mathcal{P}_t f\rangle}_{\psi(t)}.$$

Gronwall's inequality implies that

$$\mathrm{Var}(\mathcal{P}_t f) = \psi(t) \le e^{-2\lambda_1 t}\psi(0) = e^{-2\lambda_1 t}\mathrm{Var}(f). \qquad \square$$

Now, how does one make sense of time reversibility? Let X_t be a nice diffusion

$$dX_t = b(t, X_t)\, dt + \sigma(t, X_t)\, dB_t$$

which has density p_t. Let $h \in \mathcal{C}^{1,2}$ with $(\partial_t + \mathcal{L})h = 0$ be strictly positive. $Z_t = h(t, X_t)$ is a local martingale.

We will normalize h as follows. $h(0, X_0) = 1$ implying that $Q_t = h(t, X_t)\mathcal{P}_t$. Suppose that $\mathbb{E}h < \infty$. Consider $h^Y(s,x) = \frac{p(s,x,T,y)}{p(0,x_0,T,y)}$. By Markov property,

$$\mathbb{E}(Z^Y|X_s) = \int h^Y(t,x')p(s,X_s,t,x')\, dx' = \frac{p(s,X_s,T,y)}{p(0,x_0,T,y)}$$

so it is a martingale. For the kernel Q^Y, one needs

$$p(A) = \mathbb{E}\left[\mathbb{P}(A|X_T)\right] = \int Q^Y(A)p(0,x,T,y)\, dy.$$

Take $A \in \mathcal{F}_s$ for $s \le T$. Then

$$\mathbb{E}_{x_0}\left[\mathbb{1}_A g(X_T)\right] = \mathbb{E}\left[\mathbb{1}_A \mathbb{E}\left[g(X_T)|X_s\right]\right] = \mathbb{E}\left[\mathbb{1}_A \int g(y)p(s,X_s,T,y)\, dy\right]$$

so $\mathbb{P}_x(A|X_T) = q(X_T)$ and

$$q(y) = \mathbb{E}\left[\mathbb{1}_A \frac{p(s, X_s, T, y)}{p(0, x_0, T, y)}\right].$$

And so,

$$Q^Y(A) = \mathbb{E}\left[\mathbb{1}_A \frac{p(s, X_s, T, y)}{p(0, x_0, T, y)}\right].$$

Let

$$d\tilde{X}_t = \left[b(t, \hat{X}_t) + \sigma\sigma^T \nabla \log h(t, \hat{X}_t)\right] dt + \sigma(t, \hat{X}_t)\, d\tilde{B}_t$$

where $\tilde{B}_t$ is a Q-Brownian motion. The idea is that $h(t, X_t) = \exp[\log h^Y(t, X_t)]$ and

$$d \log h = -\frac{\left\|\sigma^T h\right\|^2}{2h^2}\, dt + \sigma^T \nabla \log h(t, X_t)\, dB_t.$$

Girsanov's theorem gives

$$\tilde{B} = B - \int \sigma^T \nabla \log h\, dt$$

which is a Q-Brownian motion. Therefore,

$$dX_t = b\, dt + \sigma\, dB = \left(b + \sigma\sigma^T \nabla \log h\right)\, dt + \sigma\, d\tilde{B}.$$

12 Schilder's Theorem

Let X_t^ε solve

$$\begin{cases} dX_t^\varepsilon = \sqrt{\varepsilon}\, dB_t + b(X_t^\varepsilon)\, dt \\ X_0 = 0. \end{cases}$$

We know that $X^\varepsilon \to X$ weakly with

$$\begin{cases} \dot{X} = b(X) \\ X_0 = 0. \end{cases}$$

As $\varepsilon \to 0$, the question we want to answer is how unlikely that

$$\mathbb{P}\left(\|X^\varepsilon - X\|_{\mathcal{C}([0,1],\mathbb{R}^d)} \geq \delta\right)$$

Let $\gamma_t \in \mathcal{C}([0,1], \mathbb{R}^d)$ and consider $\mathbb{P}(X^\varepsilon \in b_\delta(\gamma))$.

Definition 110. *The b-action of a trajectory is*

$$I(f) = \int_0^1 |\dot{f}(t) - b(f(t))|^2\, dt.$$

Theorem 111. *(Informal version of Schilder's theorem)*
$\mathbb{P}(X^\varepsilon \in b_\delta(\gamma)) \approx e^{\frac{-1}{\varepsilon} I(\gamma)}$.

Let $\mathcal{X}$ be a complete separable metric space.

Definition 112. *A function f is called lower semicontinuous if for all k, $\{f \le k\}$ is closed.*

Definition 113. *We say that a sequence $(\mu_n) \subseteq \mathcal{M}_1(\mathcal{X})$ admits a large deviation principle with speed a_n and rate function I if*

1. *$I : \mathcal{X} \to [0, \infty]$ is lower semicontinuous,*
2. *For all measurable E,*

$$-\inf_{x \in int(E)} I(x) \le \underline{\lim} \frac{1}{a_n} \log \mu_n(E) \le \overline{\lim} \frac{1}{a_n} \log \mu_n(E) \le -\inf_{x \in \overline{E}} I(x).$$

Let (Y_ℓ) be iid random variables with

$$\Lambda(\lambda) = \log \mathbb{E} e^{\lambda Y} < \infty$$

for all $\lambda \in \mathbb{R}$ and

$$\Lambda^*(y) = \sup_\lambda [\lambda y - \Lambda(\lambda)]$$

Theorem 114. *(Cramer)*
Denote $M_n = \frac{1}{n} \sum_{\ell=1}^n Y_\ell$ with laws $\mathbb{P}_{M_n}$. Then $(\mathbb{P}_{M_n})$ have a large deviation principle with speed n and rate function Λ^*.

Lemma 115. *(Varadhan's lemma)*
Let $\mathbb{P}_n$ have large deviation principle with speed a_n, rate function I, and F be bounded continuous function. Then

$$\lim \frac{1}{a_n} \log \mathbb{E} e^{a_n F(X_n)} = \sup_x [F(x) - I(x)]$$

where $X_n \sim \mathbb{P}_n$.

Lemma 116. *(Contraction principle)*
Suppose that $f : X \to Y$ is continuous where X, Y are complete separable. Denote $Q_n(A) = f_* \mathbb{P}_n(A) = \mathbb{P}(f(X_n) \in A)$ then Q_n has a large deviation principle with speed a_n and rate function $J(y) = \inf_{x \in f^{-1}\{y\}} I(x)$.

Theorem 117. *Let Q^ε be the law of $\sqrt{\varepsilon} B$ then it has a large deviation principle with speed $\frac{1}{\varepsilon}$ and rate function $I(\gamma) = \int_0^1 |\dot{\gamma}|^2 \, dt$.*

Theorem 118. *(Schilder)*
Let $\mathbb{P}^\varepsilon$ be the law of X^ε then $(\mathbb{P}^\varepsilon)$ has a large deviation principle with speed $\frac{1}{\varepsilon}$ and proper rate function I_b.

Using 117, one can prove 118 as follows. Consider $\Phi : \mathcal{C}[0,1] \to \mathcal{C}[0,1]$ by $g \mapsto g(t) + \int_0^t b(g(s)) \, ds$. Gronwall's inequality shows that Φ is Lipschitz. And so, $\mathbb{P}^\varepsilon$ has large deviation principle with speed $\frac{1}{\varepsilon}$ and rate function

$$J_b(\gamma) = \inf_{\Phi(g) = \gamma} \int |\dot{g}|^2 \, dt = \int |\dot{\gamma} - b(\gamma)|^2 \, dt.$$